HOW DO PLANTS ACCOMMODATE IMMUNE SURVEILLANCE AND COMMENSAL COMMUNITY FORMATION?

Nicholas R. Colaianni

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Bioinformatics and Computational Biology

Chapel Hill
2021

Approved by:

Jeffery L. Dangl

Corbin D. Jones

Terrence S. Furey

Youssef Belkhadir

Zachary L. Nimchuk

ABSTRACT

Nicholas R. Colaianni: How do plants accommodate immune surveillance and commensal community formation?
(Under the direction of Jeffery L. Dangl and Corbin D. Jones)

Every plant cell is equipped with the machinery capable of identifying, responding to, and communicating the presence of microbial invaders. Plant cells use both intracellular and extracellular receptors to recognize the presence of microbial derived molecules. Investigations into which molecules these receptors surveil for has mainly focused on a small number of pathogen-derived molecules that induce an immune response capable of reducing bacterial growth. However, the advancement and increased usage of genome sequencing technologies has revealed a substantial diversity of microbially-derived molecules. Additionally, data suggests that many microbes present in plant-associated communities contain molecules that likely bind to plant receptors and induce an immune response. These commensal communities do not induce macroscale defense phenotypes indicating that defense is either off, slightly activated, or targeted. Thus, to understand how plants form commensal communities, how much diversity exists in defense inducing molecules, how this diversity arises, what responses this diversity induces, and how plant-associated communities persist when immune inducing microbial derived molecules are present are crucial unsolved questions.

My collaborators and I leveraged advancements in high throughput interaction and functional assays, an abundance of high-quality microbial genomes isolated from a single host, and host-relevant synthetic bacterial communities to better understand the interplay between the

immune system and commensal community formation. We find that bacteria employ at least three mechanisms to produce diverse ligands that reduce immune activation and improve colonization. We present evidence that plants monitor and adjust immune responses to the proportion of these diverse functional ligands to allow the formation of communities while monitoring for pathogens. We also show that commensal communities contain many, non-overlapping suppression mechanisms that can influence the plant immune response. Collectively, these data revealed new responses to defense inducing molecules and how commensal bacteria influence defense activation. It also suggests that the plant immune system acts as a community sensor that surveils for host dysbiosis that is a predictor of pathogen infection. Future work will dissect if and how this surveillance system promotes and maintains diversity in commensal communities. This will improve our ability to manipulate plant communities and harness the full power of plant growth-promoting microbial amendments.

To my family, friends, and mentors, I am forever grateful for your support, belief in me, guidance, and friendship. This accomplishment would not have been possible without you all.

ACKNOWLEDGEMENTS

There are many people to thank for helping me throughout this process. I first want to thank my mentors Jeff Dangl and Corbin Jones. Both of you were instrumental to my success and growth as a scientist. Jeff, I want to thank both you and Sarah for creating an intellectually stimulating and collaborative lab atmosphere. I learned so much from both of you and everyone in the lab. You helped keep me focused on the task at hand, while still listening and challenging my crazy ideas and thoughts. I thoroughly enjoyed our meetings and talks over the years. It has made me a better scientist. Corbin, you were one of the main reasons I got into bioinformatics. You inspired me to learn as much as I can and were always there for me when I had issues and concerns. I am indebted to you for allowing me to work in your lab 6 years ago and taking the time to mentor me over this time period.

I also want to thank all of the amazing people in the Dangl lab I've worked with over the years. I have learned so much from each and every one of you. You all made me strive to become a better scientist and graciously showed me the way. You all picked me up when I was feeling down and were also there to celebrate with me during the good times. I will miss the conversations over coffee, during lunch, and after lab meetings.

I want to give a special thanks to my collaborators Youssef Belkhadir and Kate Parys. I will never forget both of your kindness while hosting me in Vienna. I learned so much during my trip and enjoyed working with the whole Belkhadir lab. I am also grateful to my committee, Jeff Dangl, Corbin Jones, Terry Furey, Youssef Belkhadir, and Zach Nimchuk, for taking the time to help me out through the PhD process. I have enjoyed working and speaking with each of you.

Lastly, I want to thank my family for their support. While you didn't understand what I was doing, you were always rooting for my success and willing to listen to my problems. I want to also thank my loving partner, Lindsay Wickersham. You have been with me every step of the way and have always been so supportive. I am so thankful to have you in my life.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

An	Antagonist
An-P	Antagonist Primer
AP	Antagonistic Pleiotropy
AUC	Area Under the Curve
BAK1	Bri-Associated Kinase 1
BIK1	Botrytis-Induced Kinase 1
Co-IP	Co-immunoprecipitation
Csp22	22-amino acid region of the cold shock protein
CYP71A12	*Cytochrome P450 71A12*
ECD	Extracellular Domain
Elf18	18-amino acid region of the elongation factor Tu protein
EF-Tu	Elongation factor Tu
EFR	EF-Tu Receptor
ETI	Effector-triggered immunity
EV	Empty Vector
FDR	False Discovery Rate
Flg22	22-amino acid region within FliC
FliC	The most abundant protein of the bacterial flagellin
FLS2	Flagellin sensing locus 2
FRK1	Flg22-induced receptor-like kinase 1
GCI	Grated-Coupled Interferometry
HMM	Hidden Markov Model

IQR	Interquartile range
LRR	Leucine-rich repeats
MAMP	Microbe associated molecular pattern
MIT	Mutual Information Theory
MTI	MAMP-triggered immunity
NB	No Bacteria
niSEI	non-immunogenic significantly enhanced interaction
Nlp20	20-amino acid region of the _____
NLR	Nucleotide-binding Leucine-rich Repeat
OTU	Operational Taxonomical Unit
PCA	Principal Component Analysis
PRR	Pattern Recognition Receptors
RGI	Root Growth Inhibition
ROS	Reactive Oxygen Species
SD	Standard Deviation
SEI	Significantly Enhanced Interaction
SGI	Seedling Growth Inhibition
SynCom	Synthetic Community
T2SS	Type II Secretion System
T3SS	Type III Secretion System
TLR5	Toll-like receptor 5
Useq	Unique V3-V4 16S sequence variant

CHAPTER 1: THE INTERSECTION BETWEEN PLANT IMMUNITY AND MICROBIOTA COMMUNITY FORMATION[1]

Section 1.1: Introduction

Scientists have studied the molecular aspects of plant immunity since the mid-1980s. Investigation of how plants recognize, respond, and limit the growth of invading organisms, and the strategies used by pathogens to counteract plant immunity led to a conceptual framework of the plant immune system [1]. Briefly, plants possess receptors that recognize non-self or modified-self molecules which indicate the presence of potential invaders. A first layer of pattern recognition receptors (PRRs) located in the plasma membrane perceives the presence of extracellular molecules, which are often conserved across whole classes of microbes (e.g., fungal chitin or bacterial flagellin) and are thus known as Microbe-Associated Molecular Patterns (MAMPs). Recognition of MAMPs leads to an immune response known as MAMP-triggered immunity (MTI), which is sufficient to halt the proliferation of most microbes. However, adapted pathogens have evolved effector molecules to interfere with MTI and host physiology. The clear dichotomy between extracellular MAMPS and intracellular effectors is, however, increasingly blurred [2]. In turn, plants deploy a second level of receptors to counteract adapted pathogens. These receptors belong to the family of NLR proteins (Nucleotide-binding Leucine-rich Repeat)

[1] This chapter previously appeared as part of a review article in the Current Opinion of Microbiology. The original citation is as follows: Paulo José PL Teixeira*, **Nicholas R Colaianni***, Connor R Fitzpatrick, Jeffery L Dangl, Beyond pathogens: microbiota interactions with the plant immune system, Current Opinion in Microbiology, Volume 49, 2019, Pages 7-17, ISSN 1369-5274, https://doi.org/10.1016/j.mib.2019.08.003. * indicates that these authors contributed equally to the work.

and function as intracellular sensors that recognize the presence of specific effector proteins. Direct or indirect perception of pathogen effectors by a correspondingly specific host NLR protein activates the Effector-triggered immunity (ETI), which is a robust disease-resistance response that often includes localized host cell death and systemic defense signaling. Complex interplay between plant hormones that control defense versus growth trade-offs are a major part of the plant immune system [3].

Though this model provides a good overview of the fundamental principles governing plant immunity, it is based on the interaction of plants with pathogenic microbes. However, it is clear that plants establish intimate relationships with diverse commensal microorganisms, forming complex communities in both above- and below-ground tissues (i.e. microbiomes), which vary across host plants and environments. In fact, most microbes with which plants interact are non-pathogenic [4], yet many of them express molecules that are potentially recognized by the plant immune system. Thus, one major question of plant microbiome research is whether and how the plant immune system distinguishes commensals from pathogens during microbiome assembly.

Section 1.2: Evidence for the participation of the plant immune system in microbiota assembly

Plants host microbiota whose composition differs from the surrounding environment [5-8], yet the mechanisms governing the recruitment of these microbes are largely unknown. The holobiont framework proposes that plants and their collective microbiota form a single entity subject to evolutionary processes [9,10], which implies that plants have adapted ways to distinguish their evolutionary partners from other microorganisms. Alternatively, the assembly of at least some of the plant microbiome may represent mere niche filling, a process influenced by plant traits but of minor adaptive importance for the plant host. Regardless of the proposed

adaptive value of the plant microbiome as a whole, it is obvious that some fraction of the commensal community has adaptive value to the host. It is also commonly hypothesized that plants can distinguish between pathogenic and commensal microbes. In fact, in an analysis of closely-related commensal and pathogenic *Pseudomonas* strains, the transition between lifestyles is based on the gain/loss of only a very few virulence islands [11]. Additionally, inspection of 627 bacterial genomes derived from healthy Arabidopsis roots and leaves revealed that 608 bacteria (97%) have the potential to produce putatively immunogenic MAMPs (Figure 1.1A). Many of these bacteria share identical MAMP variants to known pathogens, indicating that the MTI response shown to repress the growth of pathogens [12] is potentially activated in response to commensals as well. This begs the question of how commensals avoid or suppress MTI. It is plausible that MTI can both inhibit pathogens and maintain microbiome homeostasis by gating microbes 'in' or 'out' upon intimate contact with the host. In this view, MTI functions as a general mechanism used by plants to control the assembly of their microbiota. In support, Arabidopsis multi-mutants defective in MAMP recognition and downstream MTI signaling exhibit reduced defense against an avirulent mutant of a normally pathogenic *Pseudomonas syringae* strain and are unable to maintain normal leaf endophytic bacterial communities under high humidity. Inability to regulate the growth of endophytic bacterial communities led to mild chlorosis and necrosis in some leaves, resembling dysbiosis [13]. This indicates that plants control the growth of microbial populations with their immune system in order to maintain their own health.

Section 1.3: Strategies used by microbiota members to evade or suppress plant immunity

Plant immune receptors do not distinguish between microbial lifestyles and recognize ligands that can be present in both pathogens and commensals [14]. Evasion or suppression of

host immune responses is a hallmark of successful pathogens. Likewise, colonization by individual members of the plant microbiome, the essence of community assembly, likely requires strategies to avoid or interfere with plant immunity [15]. Recent work highlights differences and similarities between commensals and pathogens in the strategies used to suppress or evade the plant immune system. MTI suppression is used by pathogenic microbes to bypass the plant immune system [1] but has also recently been reported for non-pathogenic microbes [16-19]. The beneficial rhizobacterium *Pseudomonas simiae* WCS417 promotes plant growth and suppresses part of the transcriptional response that is triggered by the bacterial MAMP flg22 [14,20]. Similarly, specific Rhizobiales strains that colonize Arabidopsis roots are able to prevent responses that are triggered by the same MAMP [21]. Endophytes can also prevent MAMP-triggered cytosolic calcium influx in Arabidopsis [22]. A recent study found that the plant growth-promoting bacterium *Pseudomonas capeferrum* WCS358 produces organic acids that lower the extracellular pH and interfere with the response to flg22 [23]. Recent research also indicates that mutualistic fungi gain access to plant tissues by manipulating innate plant immunity [24]. Yet, the mechanisms involved in the suppression of immune responses by commensals and mutualists are still largely unexplored. The type III secretion system (T3SS) is a common feature among pathogenic bacteria and it can also be found in non-pathogenic strains [25-27]. Nevertheless, genes encoding this effector-delivery machinery are rare in the genomes of plant-associated commensal bacteria [28]. This may reflect the apparently weak host specialization of most plant-associated commensals [29]. Thus, a diversity of alternative strategies to interfere with the host immune responses, particularly on the extracellular battleground of MTI, are expected to be found in plant associated microbiota.

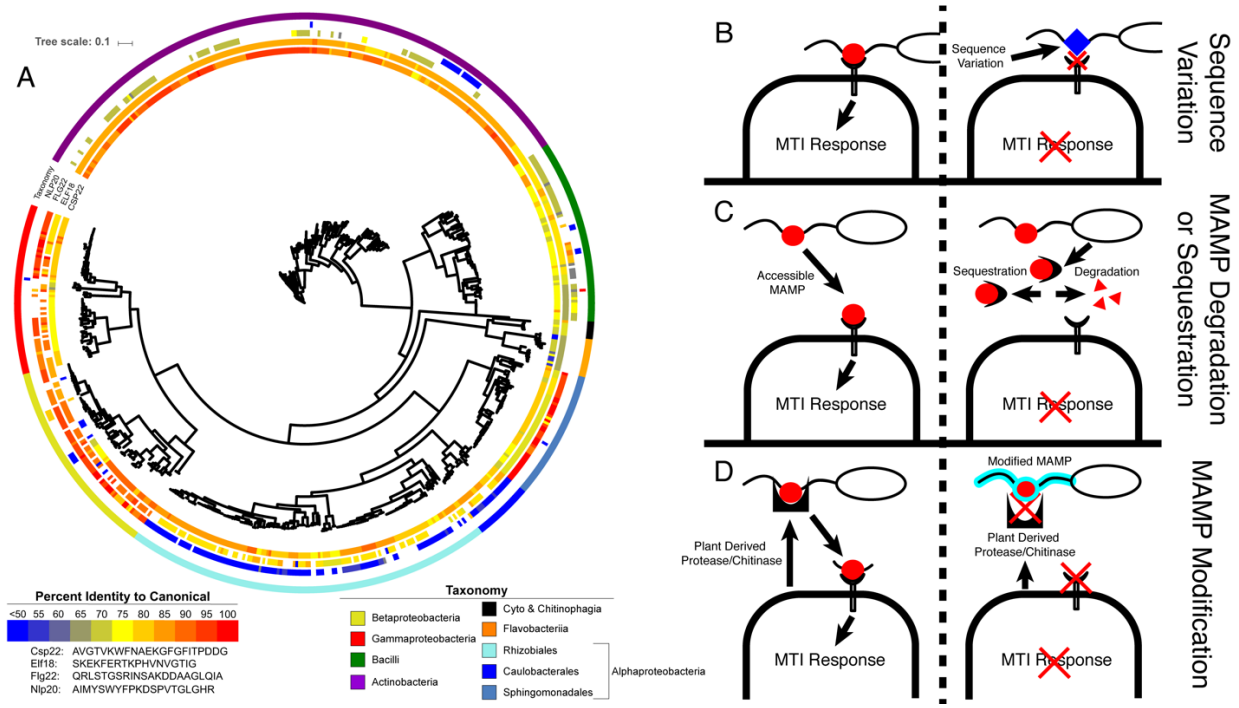


Figure 1.1. Commensal MAMP profiles and potential mechanism used by microorganisms to evade the plant immune system. A) Bacteria isolated from healthy Arabidopsis plants harbor potentially immunogenic MAMPs while others show large divergence from the canonical, which may contribute to evasion of the plant immune system. The tree includes 627 Arabidopsis-derived isolates and was generated by pruning the 3837 microbial tree in Levy et al. [28]. MAMP genes were identified using custom built profile hidden Markov models and the MAMPs were identified after aligning all MAMP genes with MUSCLE [30,31]. Percent identity was calculated using edit distance. The tree was generated with iTOL [32]. The canonical sequences are from *Micrococcus lysodeikticus* (csp22) [33], *Escherichia coli* (elf18) [34], *Pseudomonas syringae pv tabaci* (flg22) [35], and *Phytophthora parasitica* (nlp20) [36]. MAMP divergence could lead to plant immune evasion by B) sequence variation, C) MAMP degradation or sequestration, or D) MAMP modification as detailed in the text.

Section 1.4: MTI evasion mechanisms

I) MAMP divergence

Microbes might evade MTI by evolving MAMP variants that no longer bind to or activate the corresponding plant PRR (Figure 1.1B). This evasion is at face value likely to be counter-adaptive, since alteration of MAMP sequences and structures may impair the positive function of the microbial MAMP-containing molecule. For example, some flg22 variants that lose immunogenicity also lose motility [37]. Nevertheless, diverse, potentially immune evasive

MAMP variants are widespread in certain bacterial taxa. This distribution is likely to be MAMP-dependent (Figure 1.1A): 26% of the flg22 peptide epitopes found in 125 Arabidopsis-associated bacterial isolates have at least 50% sequence divergence from the canonical active sequence, while less than 1% of elf18 variants identified diverge from the canonical epitope by at least 50%. This suggests that flg22 recognition imposes stronger bacterial fitness defects and/or that the flg22 region is more amenable to variation than elf18, and/or that there are differential stringencies in the requirements for MAMP recognition. Consistent with the first hypothesis, the elf18 receptor EFR is not expressed in Arabidopsis roots, while FLS2 is [20,38]. The MTI response produced by MAMP sequences divergent from the respective canonical sequences is still relatively unknown. In parallel to the MAMP sequence diversity found across plant-associated bacteria, Arabidopsis and tomato lines display large variation in their response to different MAMPs and even to the same MAMP variant, indicating that MAMP recognition across plant populations is evolving [39-41]. Thus, recognition may be driven by the MAMP repertoire of local commensal and beneficial microbes, not just pathogens [14,21,35]. Tolerance for the former must be balanced by intolerance of the latter.

II) MAMP degradation/sequestration

Even if microbes express an immunogenic MAMP, they may have mechanisms to evade MTI. Microbes have evolved proteases that digest their MAMPs or proteins that sequester MAMPs to hide them from plant receptors (Figure 1.1B). The plant pathogen *P. syringae* DC3000 secretes a protease that, through the degradation of flagellin, decreased MTI and increased this strain's growth in Arabidopsis and tomato leaves [42]. Chitin, a conserved component of the cell wall in fungi, is another potent inducer of MTI. Several fungal pathogens have evolved chitin-binding proteins (LysM or inactive chitinases) capable of sequestering free

chitin fragments to prevent the activation of plant PRRs [43-45]. Although this mechanism has been only demonstrated for pathogens thus far, LysM and presumably inactive chitinases are found throughout the fungal kingdom and, therefore, commensals may apply this method of evasion as well.

III) MAMP modification

Another strategy used by microbes to prevent the elicitation of MTI is MAMP modification (Figure 1.1C). For example, *Nicotiana benthamiana* secretes glycosidases that strip the glycan shield from the bacterial flagellum, allowing plant proteases to release the immunogenic flg22 peptide for recognition by the FLS2 receptor [46]. In turn, pathogens can evade flg22 recognition by either inhibiting plant glycosidases or by modifying the glycan moieties that cover their flagellum [46]. Similarly, fungi can escape the plant immune system by deacetylating the chitin in their cell wall into chitosan, which is a weaker inducer of immunity [44]. Because both flg22 and chitin are ubiquitous in plant microbiomes, it is likely that commensals have evolved analogous MAMP modification strategies to evade MTI.

Section 1.5: How MAMP variability arises

While microbial genes encoding MAMPs have adaptive functions and are thus evolutionarily conserved, it is clear that diverse MAMP variants can be produced by bacterial commensals. This indicates that host immune responses impose evolutionary pressure on MAMP containing proteins that is strong enough to favor mutations in key functional regions. This interplay between the evolutionary pressure from gene function and the host immune response drives antagonistic pleiotropy (AP). The role AP plays in balancing the functionality of MAMP-coding genes against their immunogenicity is unknown and may provide mutational strategies to evolve non-immunogenic MAMP variants. To address this, we focused on an epitope of flagellin

that triggers antibacterial immunity in plants. Flagellin is conserved because it enables motility. Here, we decode the immunogenic and motility profiles of this flagellin epitope and determine the spectrum of amino acid mutations that drives AP. We discover two synthetic mutational tracks that undermine the detection activities of a plant flagellin receptor. These tracks generate epitopes with either antagonist or weaker agonist activities. Finally, we find signatures of these tracks layered atop each other in natural *Pseudomonads*.

Section 1.6: Characterization of flg22 sequence and response variability found in commensal communities

The identification of substantial flg22 diversity and mutational strategies employed to produce non-immunogenic or antagonistic flg22 variants made it unclear how immune responses are tuned to or by MAMP diversity present in commensal microbiota. We systematically studied the variability of commensal peptide derivatives of flagellin (flg22). Most flg22 peptides evade recognition, while others contribute to evasion by manipulating immunity through antagonism and signal modulation. We define substantial functional diversity and establish a paradigm of signal integration, wherein the sequential signaling outputs of the flagellin receptor are separable and allow for reprogramming by commensal-derived flg22 epitope variants. Within plant-associated synthetic communities immune evading flg22 epitopes are enriched, but upon physiological stress that represses the immune system, immune-activating flg22 epitopes become enriched. The existence of immune-manipulating epitopes suggests that they evolved to either communicate or utilize the immune system for host colonization and can influence commensal microbiota community composition. We propose that for the plant to balance immune function and allow commensal community formation, the plant immune system must work as a molecular

tabulator that responds to the changes in the proportion of flg22 variants present in the commensal community.

Section 1.7: Are commensal communities able to suppress immunity induced by flg22

It is likely necessary for commensal bacterial communities to dampen immune responses produced by the presence of immunogenic flg22 variants produced by other commensals. The importance of MAMP detection by the plant immune system is underscored by the large diversity of strategies employed by pathogens to interfere with MTI and by the fact that failure to do so is often associated with loss of virulence. Yet, whether or how MTI functions beyond pathogenic interactions is not well understood. Here we demonstrate that a community of root commensal bacteria modulates a specific and evolutionarily conserved sector of the Arabidopsis immune system. We identify a set of robust, taxonomically diverse, MTI suppressor strains that are efficient root colonizers and, notably, specifically enhances the colonization capacity of other tested commensal bacteria. We highlight the importance of extracellular strategies for MTI suppression by showing that the type-2, not the type-3, secretion system is required for the immunomodulatory activity of one robust MTI suppressor. Our findings suggest that root colonization by commensals is controlled by MTI which, in turn, can be selectively modulated by specific members of a representative bacterial root microbiota.

REFERENCES

1. Jones JD, Dangl JL, The plant immune system. *Nature*. **444**, 323-329. (2006)

2. Thomma BP, Nurnberger T, Joosten MH, Of PAMPs and effectors: the blurred PTI-ETI dichotomy. *Plant Cell.* **23**, 4-15. (2011)

3. Berens ML, Berry HM, Mine A, Argueso CT, Tsuda K, Evolution of Hormone Signaling Networks in Plant Defense. *Annu Rev Phytopathol.* **55**, 401-425. (2017)

4. Vorholt JA, Microbial life in the phyllosphere. *Nat Rev Microbiology.* 10, 828-840. (2012)

5. Hacquard S, Garrido-Oter R, Gonzalez A, Spaepen S, Ackermann G, Lebeis S, McHardy AC, Dangl JL, Knight R, Ley R, et al., Microbiota and Host Nutrition across Plant and Animal Kingdoms. *Cell Host Microbe.* **17**, 603-616. (2015)

6. Bulgarelli D, Rott M, Schlaeppi K, Ver Loren van Themaat E, Ahmadinejad N, Assenza F, Rauf P, Huettel B, Reinhardt R, Schmelzer E, et al., Revealing structure and assembly cues for Arabidopsis root-inhabiting bacterial microbiota. *Nature.* **488**, 91-95. (2012)

7. Bai Y, Muller DB, Srinivas G, Garrido-Oter R, Potthoff E, Rott M, Dombrowski N, Munch PC, 362 Spaepen S, Remus-Emsermann M, et al., Functional overlap of the Arabidopsis leaf and root microbiota. *Nature.* **528**, 364-369. (2015)

8. Lundberg DS, Lebeis SL, Paredes SH, Yourstone S, Gehring J, Malfatti S, Tremblay J, Engelbrektson A, Kunin V, Del Rio TG, et al., Defining the core *Arabidopsis thaliana* root microbiome. *Nature.* 488, 86-90. (2012)

9. Vandenkoornhuyse P, Quaiser A, Duhamel M, Le Van A, Dufresne A, The importance of the microbiome of the plant holobiont. *New Phytol.* **206**, 1196-1206. (2015)

10. Hassani MA, Duran P, Hacquard S, Microbial interactions within the plant holobiont. *Microbiome.* **6**, 58. (2018)

11. Melnyk RA, Hossain SS, Haney CH, Convergent gain and loss of genomic islands drive lifestyle changes in plant-associated *Pseudomonas*. *ISME J.* **13**, 1575-1588. (2019)

12. Li X, Lin H, Zhang W, Zou Y, Zhang J, Tang X, Zhou JM, Flagellin induces innate immunity in nonhost interactions that is suppressed by *Pseudomonas syringae* effectors. *Proc Natl Acad Sci U S A.* **102**, 12990-12995. (2005)

13. Xin XF, Nomura K, Aung K, Velasquez AC, Yao J, Boutrot F, Chang JH, Zipfel C, He SY, Bacteria establish an aqueous living space in plants crucial for virulence. *Nature*. **539**, 524-529. (2016)

14. Stringlis IA, Proietti S, Hickman R, Van Verk MC, Zamioudis C, Pieterse CMJ, Root transcriptional dynamics induced by beneficial rhizobacteria and microbial immune elicitors reveal signatures of adaptation to mutualists. *Plant J.* **93**, 166-180. (2018)

15. Zamioudis C, Pieterse CM, Modulation of host immunity by beneficial microbes. *Mol Plant Microbe Interact.* **25**, 139-150. (2012)

16. Liu Z, Beskrovnaya P, Melnyk RA, Hossain SS, Khorasani S, O'Sullivan LR, Wiesmann CL, Bush J, Richard JD, Haney CH, A Genome-Wide Screen Identifies Genes in Rhizosphere-Associated *Pseudomonas* Required to Evade Plant Defenses. *MBio.* **9**. (2018)

17. Plett JM, Kemppainen M, Kale SD, Kohler A, Legue V, Brun A, Tyler BM, Pardo AG, Martin F, A secreted effector protein of *Laccaria bicolor* is required for symbiosis development. *Curr Biol.* **21**, 1197-1203. (2011)

18. Liang Y, Cao Y, Tanaka K, Thibivilliers S, Wan J, Choi J, Kang C, Qiu J, Stacey G, Nonlegumes respond to rhizobial Nod factors by suppressing the innate immune response. *Science.* **341**, 1384-1387. (2013)

19. Plett JM, Daguerre Y, Wittulsky S, Vayssieres A, Deveau A, Melton SJ, Kohler A, Morrell-Falvey JL, Brun A, Veneault-Fourrey C, et al., Effector MiSSP7 of the mutualistic fungus *Laccaria bicolor* stabilizes the Populus JAZ6 protein and represses jasmonic acid (JA) responsive genes. *Proc Natl Acad Sci U S A*. **111**, 8299-8304. (2014)

20. Millet YA, Danna CH, Clay NK, Songnuan W, Simon MD, Werck-Reichhart D, Ausubel 407 FM, Innate immune responses activated in Arabidopsis roots by microbe-associated molecular patterns. *Plant Cell.* **22**, 973-990. (2010)

21. Garrido-Oter R, Nakano RT, Dombrowski N, Ma KW, AgBiome T, McHardy AC, Schulze-Lefert P, Modular Traits of the Rhizobiales Root Microbiota and Their Evolutionary Relationship with Symbiotic Rhizobia. *Cell Host and Microbe.* **24**, 155-167 e155. (2018)

22. Lammertz M, Kuhn H, Pfeilmeier S, Malone J, Zipfel C, Kwaaitaal M, Lin NC, Kvitko BH, Panstruga R, Widely Conserved Attenuation of Plant MAMP-Induced Calcium Influx by Bacteria Depends on Multiple Virulence Factors and May Involve Desensitization of Host Pattern Recognition Receptors. *Mol Plant Microbe Interact.* **32**, 608-621. (2019)

23. Yu K, Tichelaar R, Liu Y, Savant N, Lagendijk E, Van Kuijk S, Stringlis I, Van Dijken A, Haney C, Pieterse C, Plant beneficial *Pseudomonas spp.* suppress local root immune responses by gluconic acid-mediated lowering of environmental pH. *CURRENT-BIOLOGY-D-19-00852.* (2019)

24. Shen Q, Liu Y, Naqvi NI, Fungal effectors at the crossroads of phytohormone signaling. *Curr Opin Microbiol.* **46**, 1-6. (2018)

25. Mavrodi DV, Joe A, Mavrodi OV, Hassan KA, Weller DM, Paulsen IT, Loper JE, Alfano JR, Thomashow LS, Structural and functional analysis of the type III secretion system from *Pseudomonas fluorescens* Q8r1-96. *J Bacteriol*. **193**, 177-189. (2011)

26. Stringlis IA, Zamioudis C, Berendsen RL, Bakker P, Pieterse CMJ, Type III Secretion System of Beneficial Rhizobacteria *Pseudomonas simiae* WCS417 and *Pseudomonas defensor* WCS374. *Front Microbiol.* **10**, 1631. (2019)

27. Preston GM, Bertrand N, Rainey PB, Type III secretion in plant growth-promoting *Pseudomonas fluorescens* SBW25. *Mol Microbiol.* **41**, 999-1014. (2001)

28. Levy A, Salas Gonzalez I, Mittelviefhaus M, Clingenpeel S, Herrera Paredes S, Miao J, Wang K, Devescovi G, Stillman K, Monteiro F, et al., Genomic features of bacterial adaptation to plants. *Nat Genet.* **50**, 138-150. (2017)

29. Thiergart T, Duran P, Ellis T, Garrido-Oter R, Kemen E, Roux F, Alonso-Blanco C, Agren J, Schulze-Lefert P, Hacquard S, Root microbiota assembly and adaptive differentiation among European Arabidopsis populations. *bioRxiv.* **640623**. (2019)

30. Edgar RC, MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792-1797. (2004)

31. Eddy SR, Accelerated Profile HMM Searches. *PLoS Comput Biol.* **7**, e1002195. (2011)

32. Letunic I, Bork P, Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* (2019)

33. Felix G, Boller T, Molecular sensing of bacteria in plants. The highly conserved RNA-binding motif RNP-1 of bacterial cold shock proteins is recognized as an elicitor signal in tobacco. *J Biol Chem.* **278**, 6201-6208. (2003)

34. Kunze G, Zipfel C, Robatzek S, Niehaus K, Boller T, Felix G, The N terminus of bacterial elongation factor Tu elicits innate immunity in Arabidopsis plants**.** *Plant Cell.* **16**, 3496-3507. (2004)

35. Felix G, Duran JD, Volko S, Boller T, Plants have a sensitive perception system for the most conserved domain of bacterial flagellin. *Plant J.* **18**, 265-276. (1999)

36. Bohm H, Albert I, Oome S, Raaymakers TM, Van den Ackerveken G, Nurnberger T, A conserved peptide pattern from a widespread microbial virulence factor triggers pattern-induced immunity in Arabidopsis. *PLoS Pathog.* **10**, e1004491. (2014)

37. Naito K, Taguchi F, Suzuki T, Inagaki Y, Toyoda K, Shiraishi T, Ichinose Y, Amino acid sequence of bacterial microbe-associated molecular pattern flg22 is required for virulence. *Mol Plant Microbe Interact.* **21**, 1165-1174. (2008)

38. Wyrsch I, Dominguez-Ferreras A, Geldner N, Boller T, Tissue-specific FLAGELLIN-SENSING 2 (FLS2) expression in roots restores immune responses in Arabidopsis fls2 mutants. *New Phytol.* **206**, 774-784. (2015)

39. Vetter M, Karasov TL, Bergelson J, Differentiation between MAMP Triggered Defenses in *Arabidopsis thaliana*. *PLoS Genet.* **12**, e1006068. (2016)

40. Veluchamy S, Hind SR, Dunham DM, Martin GB, Panthee DR, Natural variation for responsiveness to flg22, flgII-28, and csp22 and *Pseudomonas syringae* pv. tomato in heirloom tomatoes. *PLoS One.* **9**, e106119. (2014)

41. Roberts R, Mainiero S, Powell AF, Liu AE, Shi K, Hind SR, Strickler SR, Collmer A, Martin GB, Natural variation for unusual host responses and flagellin-mediated immunity against *Pseudomonas syringae* in genetically diverse tomato accessions. *New Phytol.* (2019)

42. Pel MJ, van Dijken AJ, Bardoel BW, Seidl MF, van der Ent S, van Strijp JA, Pieterse CM. *Pseudomonas syringae* evades host immunity by degrading flagellin monomers with alkaline protease AprA. *Mol Plant Microbe Interact.* **27**, 603-610. (2014)

43. Fiorin GL, Sanchez-Vallet A, Thomazella DPT, do Prado PFV, do Nascimento LC, Figueira AVO, Thomma B, Pereira GAG, Teixeira P, Suppression of Plant Immunity by Fungal Chitinase-like Effectors. *Curr Biol,* **28**, 3023-3030 e3025. (2018)

44. Sanchez-Vallet A, Mesters JR, Thomma BP, The battle for chitin recognition in plant-microbe interactions. *FEMS Microbiol Rev.* 39, 171-183. (2015)

45. de Jonge R, van Esse HP, Kombrink A, Shinya T, Desaki Y, Bours R, van der Krol S, Shibuya N, Joosten MH, Thomma BP, Conserved fungal LysM effector Ecp6 prevents chitin-triggered immunity in plants. *Science.* **329**, 953-955. (2010)

46. Buscaill P, Chandrasekar B, Sanguankiattichai N, Kourelis J, Kaschani F, Thomas EL, Morimoto K, Kaiser M, Preston GM, Ichinose Y, et al., Glycosidase and glycan polymorphism control hydrolytic release of immunogenic flagellin peptides. *Science.* **364**. (2019)

CHAPTER 2: SIGNATURES OF ANTAGONISTIC PLEIOTROPY IN A BACTERIAL FLAGELLIN EPITOPE[2]

Section 2.1: Introduction

Innate immune systems of both plants and animals have evolved to promote the elimination of invading microbes by detecting and responding to 'non-self' molecules such as microbe-associated molecular patterns (MAMPs) (Dangl et al., 2013; Iwasaki and Medzhitov, 2015). Microbial genes encoding proteinaceous MAMPs are evolutionarily conserved, broadly distributed, and indispensable for core organismal functions (McCann et al., 2012; Mott et al., 2016). Thus, when facing the activities of a host immune system, the function of these genes mediates opposed effects on microbial fitness, a concept known as antagonistic pleiotropy (AP) (Chen and Zhang, 2020). AP arises when the negative pressure to maintain vital MAMP-coding protein function is opposed to a positive pressure acting to mediate evasion from immune system detection. AP counteracts natural selection by maintaining allele variation in population despite the availability of beneficial mutations (Venkataram et al., 2020). Yet, its effect on the functionality of MAMP-coding genes is rarely, if ever, investigated. To address this, we focused on flagellin, the main structural component of the bacterial flagellum (Macnab, 2003; Yonekura et al., 2003). The flagellin gene *fliC* is conserved and under strong evolutionary pressure because

[2] This chapter previously appeared in Cell Host & Microbe. The original citation is as follows: Katarzyna Parys*, **Nicholas R. Colaianni*,** Ho-Seok Lee*, …, Niko Geldner, Michael Hothorn, Corbin D. Jones, Jeffery L. Dangl, Youssef Belkhadir, **Signatures of antagonistic pleiotropy in a bacterial flagellin epitope**, *Cell Host & Microbe*, 2021, ISSN 1931-3128, https://doi.org/10.1016/j.chom.2021.02.008. Note that * indicates that these authors contributed equally to the work. All supplemental tables and data can be accessed online.
My contributions to this work included statistical analyses, ROS burst and root antagonism assays, mutual information theory, creation of the bacterial databases used in this study, and figure creation. I also intellectually helped in experimental design and overall data interpretation.

it plays an essential role in enabling bacterial motility (Ramos et al., 2004). Flagellar- based motility is essential for the maximal virulence of several human and plant pathogens (Rossez et al., 2015). In a classic case of evolutionary convergence, metazoans and plants have independently evolved sensors that detect flagellin as a MAMP (Fliegmann and Felix, 2016). In mammals, Toll-like receptor 5 (TLR5) detect flagellin outside cells (Iwasaki and Medzhitov, 2004). Distinguishing organizational principles of the plant flagellin signaling pathways are: (1) plants perceive flagellin at the cell surface using the membrane-bound receptors FLAGELLIN SENSING 2 and 3 (FLS2 and FLS3) (Hind et al., 2016), (2) plant flagellin sensors detect, and are activated by, short peptide epitopes that are likely derived from degradation products of flagellin, as opposed to TLR5 which senses an entire domain concealed in a folded flagellin molecule (Buscaill et al., 2019; Sun et al., 2013; Yoon et al., 2012), (3) the key signaling output, MAMP-triggered immunity, provides a rapid and broad-spectrum immune response (Hacquard et al., 2017; Lee and Belkhadir, 2020; Teixeira et al., 2019). With short microbial epitopes that physically interface with host immune receptors, the molecular mechanisms of flagellin sensing in plants provide an outstanding system to study the molecular footprints of AP.

In *Arabidopsis thaliana* (hereafter Arabidopsis), the reference plant system of this study, a 22-amino acid epitope peptide derived from the N-terminus of flagellin (flg22) is sufficient to activate FLS2 signaling (Chinchilla et al., 2006; Felix et al., 1999; Gomez-Gomez and Boller, 2000). Arabidopsis mutant plants lacking *FLS2* function are more susceptible to infection by bacterial pathogens such as *Pseudomonas syringae* (*P. syringae* or *Ps*) (Zipfel et al., 2004). In general, the genus Pseudomonas comprises a wide range of behaviorally different species that can colonize humans and/or plants (Sitaraman, 2015), and *Pseudomonads* are commonly found in the leaf microbiome of Arabidopsis (Bodenhausen et al., 2013; Chen et al., 2020; Delmotte et

al., 2009; Karasov et al., 2018). flg22 is extremely well conserved in Pseudomonas pathovars, and epitope conservation is paramount for *Pseudomonas aeruginosa* (hereafter *Pa*). Despite this extreme conservation, a few individual amino acids are polymorphic, providing some diversity in the epitope. These variations may represent bacterial attempts to evolve undetectable flg22 epitopes to avoid host recognition while remaining motile (Clarke et al., 2013). The function of flg22 on bacterial motility has been addressed in studies involving a very narrow set of mutations targeting only a couple of residues (Naito et al., 2008; Wang et al., 2015). As a consequence, we still do not understand how the evolutionary trap mediated by FLS2 drives AP, and how, in turn, this has led to the conservation or diversification of some residues in flg22.

Here, we address this knowledge gap by merging synthetic biology and directed evolution into a single approach. We used the flg22 peptide from *Pa* (Pa flg22) as a reference peptide, since its interaction with the extracellular domain of FLS2 is understood at the atomic level (Sun et al., 2013), and the molecular responses it induces are the most characterized (Chinchilla et al., 2007; Schulze et al., 2010). The central goal of the present study was to deconstruct both the immunogenic and motility profile of Pa flg22 by deep-mutational scans to determine the spectrum of amino acids in the epitope that induce AP. First, we determined the landscape of mutational impacts on motility. Next, we elucidated the principles governing FLS2-flg22 interactions and, thus, immunogenicity. We identified two distinct mutational tracks that rely on distinct mechanisms to avoid epitope detection by FLS2. One track drives a loss-of-agonist function by reducing interaction with FLS2, while the second results in gain-of-function epitopes that act as FLS2 antagonists. Finally, a computational analysis of prevailing epitopes seen in the current bacterial sequence databases indicate that the changes resulting from these synthetic tracks exist in nature.

Section 2.2: Results

2.2.1 Deep mutational scanning of flg22 reveals the spectrum of amino acids required for motility and/or interaction with FLS2

To determine the extent to which the amino acid sequence of flg22 impacts the function of flagellin, we built a library of mutant bacterial strains expressing nearly all possible single missense mutations of the flg22 domain in *Pa* (*Pa flg22*). For this, we generated 412 *fliC* allelic clones by targeting all *Pa flg22* codons for replacement by codons that specify each of the other amino acids (Figure 2.1). Next, we constructed a *Pa fliC* mutant (*Pa ΔfliC)* and transformed each individual clone of our mutant compendium into this non-motile strain to test for motility rescue (Figures 2.1A, B and 2.8A). To eliminate false negatives, we confirmed that the flagellin protein was expressed in all the transformants (Figure 2.8B). Only 18.7% of the strains displayed motility phenotypes indistinguishable from that of the wildtype *fliC* transformant (Figure 2.8C). While 45% of the transformants showed a markedly reduced motility, 34.8% remained nonmotile, similar to *Pa ΔfliC* (Figure 2.8C). Thus, the vast majority of the mutations in our deep mutational scan negatively impact motility in *Pa* (Figure 2.1C). In contrast, only a few mutations enhanced *Pa* motility (Figure 2.1D).

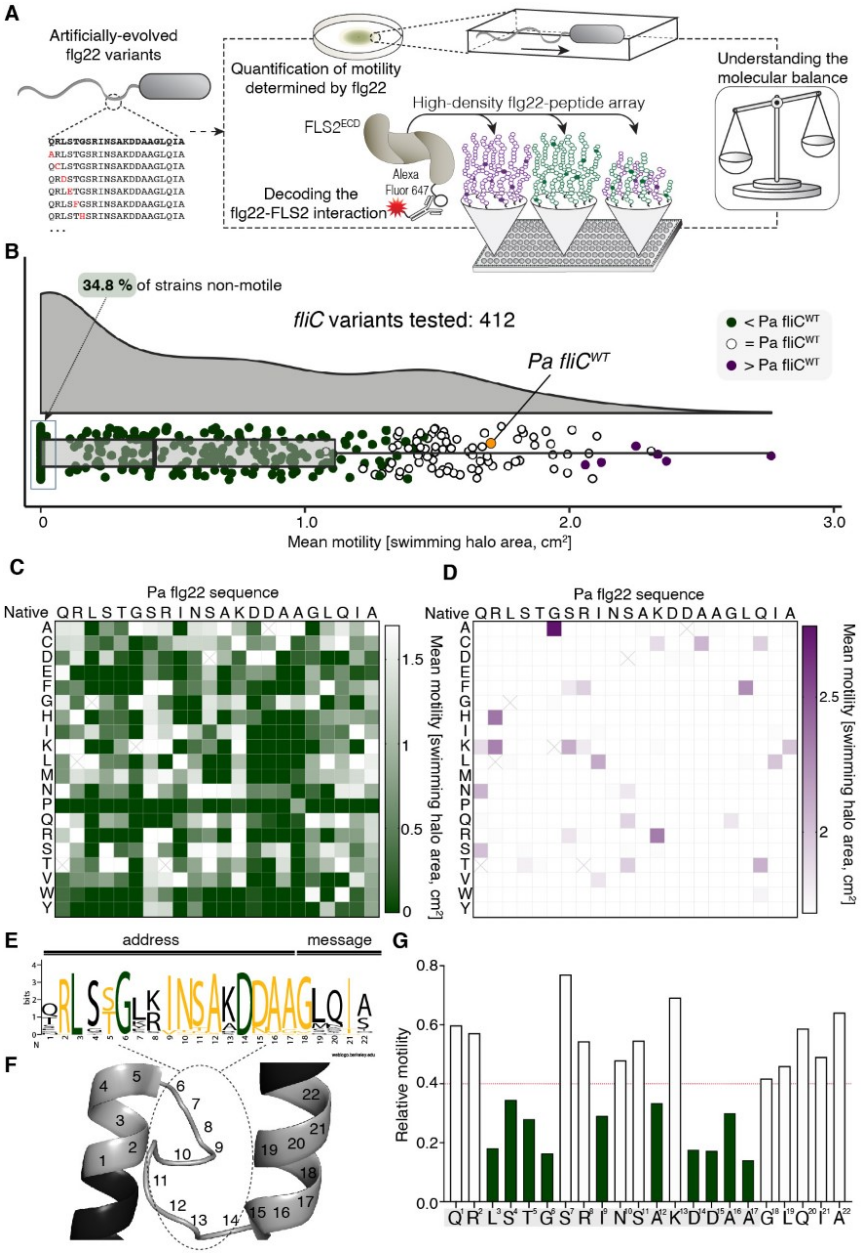


Figure 2.1. Deep mutational scans of flg22 underpin the spectrum of amino acids critical for bacterial motility. A– Mutational scanning pipelines used in the study. **B-** Raincloud plot representing the density distribution of the mean swimming motility for *Pa ΔfliC* strains complemented with allelic variants of *fliC*. The box plot shows the median motility and the quantiles. Transformants with a significantly reduced or increased motility compared to that wildtype (*fliC*^{WT}, in orange) are indicated in green and purple, respectively (Two-sided Mann-Whitney U test followed by a FDR correction; $P < 0.05$). Transformants with motilities comparable to *fliC*^{WT} are indicated in white. **C, D-** Position specific scoring matrices heatmaps representing the impact of amino acid substitutions on *Pa* motility. Reduced and increased motility is indicated in green (C) and purple (D) pixels, respectively. **E-** *Pseudomonads* flg22 sequence logo. Invariant or highly conserved (with at most two substitutions) amino acids are highlighted in green and orange, respectively. The size of the letters indicates the conservation level of each amino acid

in bits. The address/message segments and amino acids positions are indicated on top and bottom, respectively. **F-** Superimposition of flg22 (marked in grey) on flagellin structure (PDB ID: 5WK6). Amino acids numbers are indicated on the structure. **G-** Bar plots representing the relative motility (compared to wildtype) resulting from all amino acid changes at each position. The red dotted line represents half of the maximum value for the least affected position S7. Positions with a relative motility below this line are marked in green.

Based on the "address-message" concept (Meindl et al., 2000), supported and refined by the crystal structure of FLS2 in complex with Pa flg22 (Sun et al., 2013), the 22 amino acids of the epitope can be divided into two functional domains (Figure 2.1E). The 'address' segment includes the seventeen N-terminal residues and interacts with FLS2. The last five C-terminal residues define a 'message' segment that creates a docking site for the co-receptor BRI1-ASSOCIATED KINASE 1 (BAK1), and thus is important for appropriate activation of immune responses (Belkhadir et al., 2014). We used flg22 sequence conservation and structure as well as the "address-message" concept to map the impact of each amino acid changes on motility (Figures 2.1E-2.1G). We found that mutations targeting the 'message' portion of the epitope had little to no effect on motility (Figure 2.1G). Conversely, more than half of the residues in the 'address' domain (10 out of 17; >58%) were highly intolerant to mutations and are thus critical for motility (Figure 2.1G). When superimposed on the flg22 domain of the flagellin atomic structure (Wang et al., 2017), we found that these residues distributed over two small clusters that contribute structurally to a helix-loop-helix transition (Figure 2.1F). Thus, alterations of these clusters are most likely constrained by the structural requirements necessary to build a functional flagellum. The amino acids of these clusters are well conserved in flg22 epitopes derived from naturally occurring strains of *Pseudomonads* (Figure 2.1E; Table S1; Data S1), and only 1.4% of the synthetic loss-of-function mutations identified here are found in wild Pseudomonas (Table S2; Data S1).

Next, we predicted that FLS2 would focus on the conserved residues that are important for motility. To test this, we interrogated pairwise interactions between the extracellular domain of FLS2 ($FLS2^{ECD}$) and thousands of mutated flg22 peptides (Figure 2.1A). We used a high-density peptide array that comprised full substitution scans targeting both full-length Pa flg22 as well as epitope variants truncated at either or both of the N- and C-terminal moieties (Figures 2.9A-E). To test the full contribution of the 'address/message' segments to the interaction with FLS2, we additionally deleted both extremities of Pa flg22 in a stepwise manner (Figures 2.9D and 2.9E). In total, we interrogated and assigned an interaction score to over 3,000 pairwise interactions between synthetic epitopes and $FLS2^{ECD}$ (Figure 2.9A). Next, we focused on the peptides from the mutational scans of the full-length epitope and used statistics relying on interquartile range (IQR) to assign their interaction scores to distinct categories (Figure 2.2A). Peptides with an interaction score below or above the IQR of the wildtype Pa flg22 controls were considered to have weaker or stronger interactions with $FLS2^{ECD}$, respectively (Figure 2.2A). All others were classified as having a similar interaction. Peptides with an interaction score 1.5 times above or below the Pa flg22 IQR were designated as significantly different (Figure 2.2A). Using these cutoffs, 23 peptide variants showed a significantly enhanced interaction with $FLS2^{ECD}$. Only $flg22^{D14H}$ and $flg22^{D14L}$ displayed a drastically reduced interaction (Figure 2.2A). Overall, 74.5% of the mutant peptide population retained interaction with $FLS2^{ECD}$. Thus, disrupting the stability of the flg22-FLS2 interaction by single amino acid replacement is difficult. We propose that, in Arabidopsis, the stability of this interaction counteracts bacterial attempts to evade immune detection through single amino acid substitutions.

We then leveraged our dataset and mapped the individual impact of all programmed amino acid replacements on the flg22- $FLS2^{ECD}$ interaction (Figures 2.2B-2.2G). Mutations

replacing any residue of the 'message' segment with positively or negatively charged amino acids reduced or increased the flg22-FLS2^{ECD} interaction, respectively (Figures 2.2B-2.2G). Similarly, exchanging Lys^{13} for a negatively charged amino acid enhanced the interaction with FLS2^{ECD} (Figure 2.2E and 2.2F). Thus, the local electrostatic charge of both the address and message segment of the epitope can either promote or disrupt the interaction with FLS2 (Figure 2.2G). Our analyses revealed that Asp^{14} and Asp^{15} (hereafter Asp$^{14/15}$) are above all, independent of the peptide length, the most critical residues of the 'address' segment for interaction with FLS2^{ECD} (Figure 2.9F). Our findings are supported by structural and functional studies (Chinchilla et al., 2006; Sun et al., 2013). Despite the high conservation of other residues, the Arabidopsis FLS2 focuses with extreme precision on the epitope's Asp$^{14/15}$ residues, likely because this doublet is important for structurally defining a boundary of the helix-loop-helix transition in the flg22 domain of flagellin (Figure 2.1D).

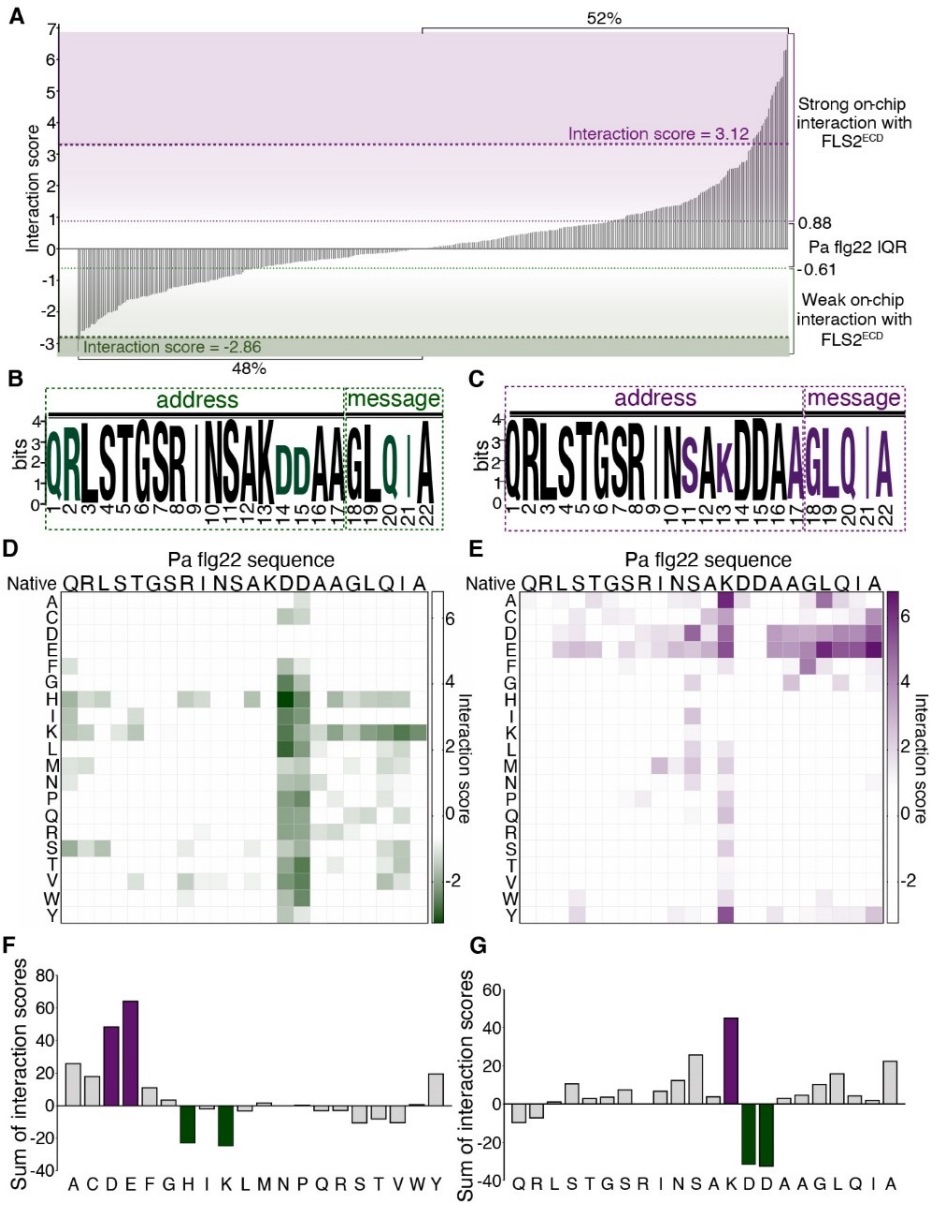


Figure 2.2. Deep mutational scans of flg22 determine the repertoire of amino acids controlling interaction with FLS2. A- Distribution of interaction scores between FLS2^{ECD} and flg22 peptides derived from the full-length mutational scans. flg22 peptides were classified as interacting with FLS2^{ECD} more strongly (purple), similarly (white), or weaker (green) based on analysis of the interquartile range (IQR). Peptides with an interaction score >3.12 or <-2.86 were designated as significantly different from all others. **B**, **C**- flg22 sequence logos mapping the FLS2 interaction scores. Amino acid mutations that reduce or enhance the strength of the flg22-FLS2^{ECD} interaction are highlighted in green or purple, respectively. Small letters indicate a more reduced (B) or improved (C) interaction. The address/message segments and amino acids positions are indicated on top and bottom, respectively. **D, E-** Position specific scoring matrices heatmaps representing the impact of each amino acid change on the flg22-FLS2^{ECD} interaction. Mutations generating flg22 variants with FLS2 interaction scores below (D) or above (E) the IQR of the wildtype Pa flg22 are represented in green or purple pixels, respectively. White pixels indicate variants with an interaction score distributing within the Pa flg22 IQR. **F-** Bar plot representing

the sum of interaction scores resulting from substitutions targeting all flg22 positions with the same amino acid. **G-** As in (F) but for all the substitutions at each amino acid position. Substitutions that strongly reduce or enhance the interaction scores are marked in green or purple, respectively.

2.2.2 Antagonistic pleiotropy constrains the evolution of non-immunogenic flg22 variants

Next, we predicted that if selective forces acted antagonistically to fine-tune bacterial motility and/or interaction with FLS2, we should be able to detect their molecular footprints. To address this, we assigned, based on the interaction score of their respective epitopes, each individual bacterial mutant strain of our collection to one of the three interaction subgroups we previously defined (Figure 2.2A). We then analyzed the motility of each group and found that the density distribution between the interaction subgroups did not differ significantly (Figure 2.3A). Then, we compared the individual contribution of each residue to both motility and interaction with FLS2 (Figure 2.10A). We found that mutations of the invariant or highly conserved residues Leu^{3}, Gly^{6}, and $Ala^{16/17}$ retained near wildtype-levels of interaction with FLS2 while having drastic consequences on motility (Figure 2.10A). A second group of evolutionary conserved residues displayed no global adverse effects on either motility maintenance or FLS2 interaction (Figures 2.1C and 2.10A). Thus, evolutionary conservation is not the sole criteria used by FLS2 to detect this flagellin epitope in Arabidopsis.

We found, however that mutations of $Asp^{14/15}$ were able to disrupt both interaction with FLS2 and bacterial motility at a higher frequency than any other residue (Figures 2.10A-C). To confirm the broader implications of these results, we tested if mutations targeting $Asp^{14/15}$ would also disrupt the motility of the host specific plant pathogen *Pseudomonas syringae* (*Ps*). We focused on $flg22^{D14K}$ and ^{D15K}, because these substitutions clearly disrupted the interaction with $FLS2^{ECD}$. First, we engineered the *Ps fliC* gene so that it encodes the Pa flg22 epitope and subsequently verified that the resulting gene complemented *Ps ΔfliC* motility (Figure 2.10D).

We then used this engineered strain to independently introduce the missense mutations targeting the $Asp^{14/15}$ doublet. Motility assays on semisolid media, as well as live imaging in liquid conditions, demonstrated that neither of these two mutant strains moved (Figure 2.3B; Video S1-S5). Next, we verified that these mutant epitopes no longer interacted with $FLS2^{ECD}$ in Grating-Coupled Interferometry (GCI) assays (Hohmann et al., 2018) (Figure 2.3C). Thus, the adaptive fixation of the $Asp^{14/15}$ doublet for the maintenance of *Pseudomonads* motility is maladaptive in the presence of FLS2 and defines a typical signature of AP. We then reasoned that bacteria must have evolved ways to alleviate the antagonistic effects mediated by these two residues. Accordingly, a few strains in our synthetic collection had a set of overlapping mutations targeting Asp^{14} or Asp^{15} that allowed partial restoration of motility while still preventing maximal interaction with FLS2 (Figure 2.10C). Since these synthetic mutations can occur as natural variations of Asp^{15} in pathogenic lineages of *Pseudomonas* that associate with Arabidopsis in the wild (Table S1 and S2; Data S1), we propose that carefully balanced trade-offs have occurred throughout evolution to maximize the escape from FLS2 detection while maintaining sub-optimal flagellin function.

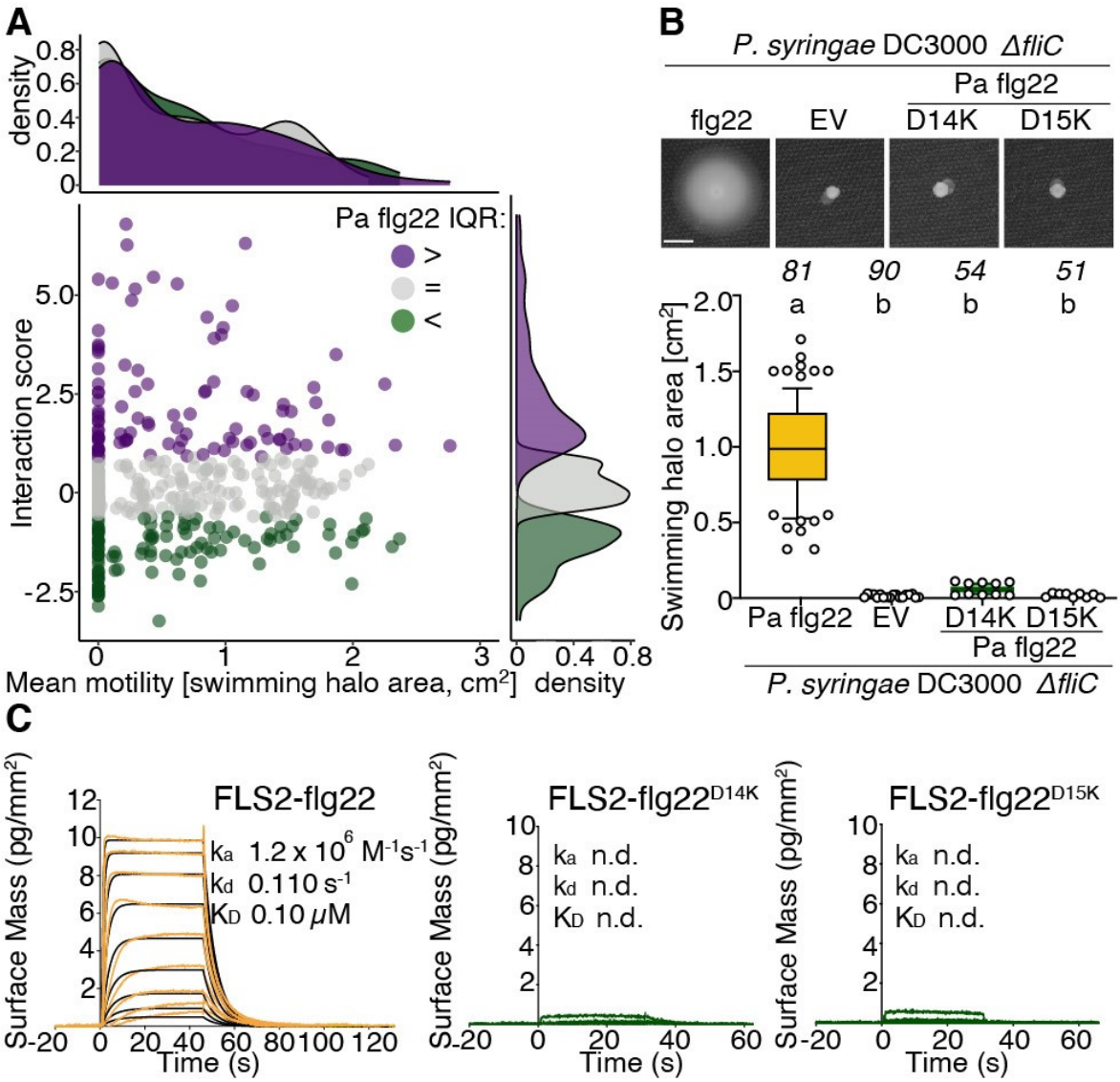


Figure 2.3. Antagonistic pleiotropy constrains the evolution of non-immunogenic flg22 epitopes. A- Scatter plot representing the relationship between the overall motility of the *Pa* transformant compendium (x-axis) and the flg22-FLS2 interaction scores (y-axis). Each transformant was assigned to an interaction subgroup based on the Pa flg22 IQR and was color-coded accordingly. The margins indicate the density distributions color-coded by interaction subgroups. The density distributions of the subgroups are not statistically different (Kolmogorov-Smirnov test, KS > 0.6). **B-** Substitution of $Asp^{14/15}$ to a Lys disrupts the swimming motility of *P. syringae*. **Top panel:** Pictures showing the motility phenotype of *Ps ΔfliC* strains complemented with an empty vector (EV) or with *fliC* variants coding either the $flg22^{D14K}$ or $flg22^{D15K}$ epitopes. Scale bar: 0.5 cm. **Bottom panel:** Quantification of swimming motility for the indicated *Ps* strains. Genotypes with the same letter are indistinguishable at >95% confidence (one-way ANOVA followed by Turkey HSD for all pairwise comparisons; $P < 0.05$). The number of biologically independent observations (n) is indicated on top of the graph. **C-** Grating-coupled interferometry (GCI)- derived interaction kinetics for $FLS2^{ECD}$ and the peptide epitopes. Shown are sensorgrams with data in orange (Pa flg22) or green ($flg22^{D14K}$/$flg22^{D15K}$), with the respective fits overlaid in black. n.d. indicates not determined.

2.2.3 Functional uncoupling of FLS2 interaction from activation as a strategy to avoid immune responses.

Next, we assessed the immunogenic functions of the mutant epitopes identified by our deep sequence/interaction mutational scans. For this, we selected, in addition to flg22^{D14K} and ^{D15K}, epitope derivatives that uniformly covered both ends of the FLS2 interaction score distribution (Figure 2.11A; Data S2 and S3). A total of 44 synthetic peptides, including ten of the variants that belonged to the group with a significantly enhanced interaction (SEI), were subsequently tested and systematically compared to the immunogenic responses of Pa flg22 in both wildtype and *fls2* mutant plants (Figures 2.4A and 2.11B-F). All our assays also included the inactive Pa flg20 peptide as an additional negative control (Felix et al., 1999). We used flg22-induced reactive oxygen species (ROS) and seedling growth inhibition (SGI) assays to measure FLS2 activation and found that immune signaling was largely contingent on interaction with FLS2 (Figures 2.4A and 2.11B-F). Most of the variants with interactions comparable to Pa flg22 induced similar immune responses, although a few elicited slightly stronger or weaker responses. Only flg22^{Q20C} had an interaction score that contrasted with its immunogenic activities and was therefore treated as an outlier (Figures 2.11G-K). We found that no epitopes from the SEI group surpassed the responses induced by Pa flg22 (Figures 2.4A and 2.11C, 2.11D and 2.11F). In fact, four of these variants displayed output responses that resembled that of flg20 (Figures 2.4A and 2.11B-F). Thus, we verified that these variants interacted retained their ability to interact with FLS2 using GCI (Figure 2.11L; Data S4).

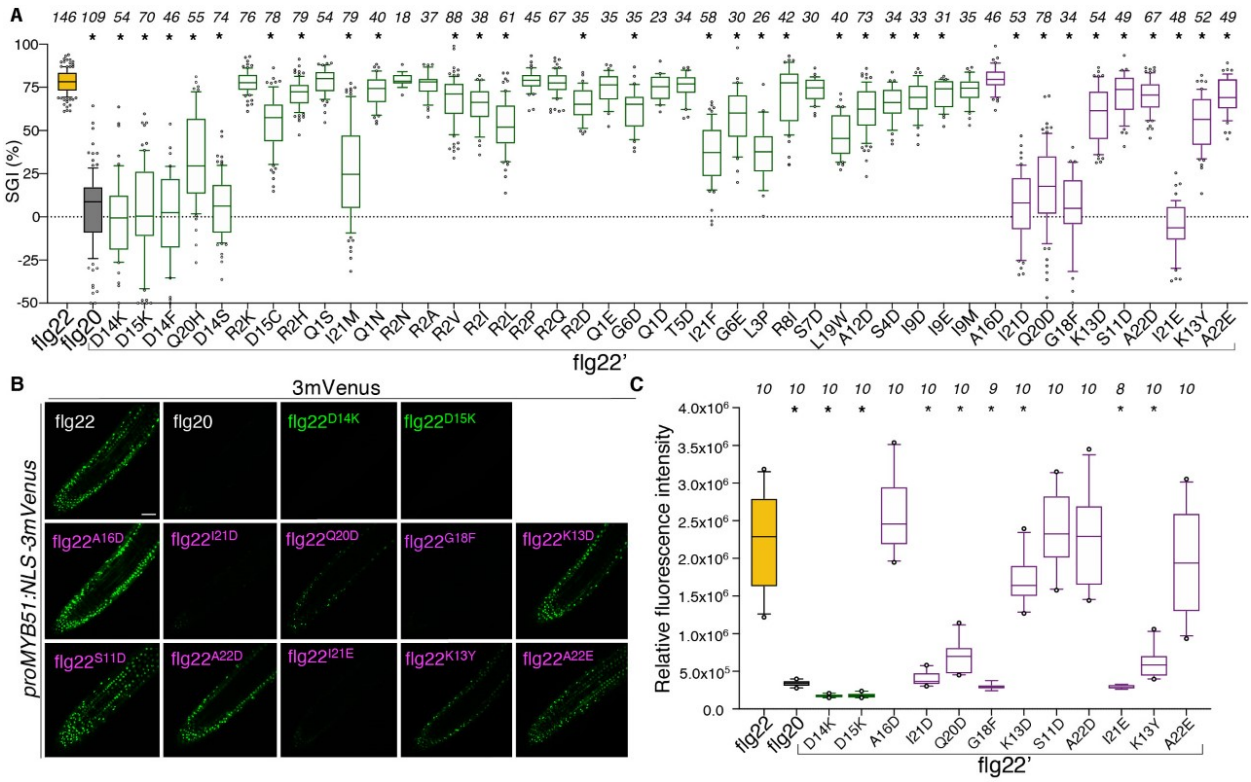


Figure 2.4. Uncoupling of FLS2 interaction from activation as a strategy to produce non-immunogenic flg22 epitopes. A- flg22-induced seedling growth inhibition (SGI). Biologically independent observations (n) is indicated on top of the graph. The flg22 peptides tested are indicated at the bottom. The peptides are ordered from left to right by increasing interaction score with FLS2. Statistical significance was assessed using a linear mixed model (two-sided *t*-test followed by multiple testing correction using the Holm method; *P < 0.05). **B-** Root expression patterns of the defense marker gene *MYB51* in response to epitope treatments. Shown are images from single confocal sections. The epitopes are indicated on top of each panel. Scale bar shown for flg22 is 50 µm and applies to all pictures. **C-** Quantitative analysis of 3mVENUS signal intensities after treatment with the flg22 peptide derivatives indicated at the bottom. The number of biologically independent observations (n) is indicated on top of the graph. * indicates statistical significance (one-way ANOVA followed by a Dunnett's multiple comparison test; P < 0.05).

To further confirm these results, we used live-imaging of immune marker gene expression and monitored FLS2 responses in the epidermal cells of cotyledons and roots using two sets of fluorescent transcriptional reporter lines (Figures 2.4B, 2.4C, 2.11M and 2.11N) (Wyrsch et al., 2015; Zhou et al., 2020). Consistent with our previous results, flg22^{D14K} and ^{D15K} were not active in these assays (Figures 2.4B and 2.4C). flg22$^{I21D,\ Q20D,\ I21E}$ and ^{G18F} proved once

more to be non-immunogenic in these assays, while all other SEI peptides displayed responses nearly equivalent to that of Pa flg22 in both cell types (Figures 2.4B, 2.4C, 2.11M and 2.11N). In contrast to flg22^{K13D} which induced a response in both root and cotyledons, flg22^{K13Y} induced a response in cotyledon cells only (Figures 2.11M and 2.11N). Thus, FLS2 responses to polymorphic epitopes are cell-type dependent in Arabidopsis. Collectively, our results show that interaction with, and activation of, FLS2 can be uncoupled. These findings are consistent with the uncoupling of FLS2 responses mediated by natural polymorphisms in commensal flg22 epitopes (Colaianni et al., submitted). We propose that this functional uncoupling could be leveraged by microbes to circumvent the pressure of AP on the Asp $^{14/15}$ residues.

2.2.4 Tunability of antagonistic pleiotropy in the message segment of flg22 licenses the evolution of non-immunogenic epitopes

Next, we assessed if the mutations defined in the group of non-immunogenic SEI peptides (hereafter niSEI) affected the motility of either *Pa* or *Ps* using the same strategy as described above. We found that strains encoding FliC variants with either the flg22^{I21D} or flg22^{G18F} epitopes restored motility of *Pa ΔfliC* to half of the levels of the unmutated FliC, while strains carrying flg22^{Q20D} and flg22^{I21E} failed to do so (Figures 2.5A and 2.12A). In contrast, these gene variants acted in opposite ways in the rescue of *Ps ΔfliC* (Figure 2.5B, 2.12B; Video S6-S9); strains encoding FliC variants carrying the flg22^{Q20D} and flg22^{I21E} epitope were able to move as well as strains carrying Pa flg22. Notably, all immunogenic variants from the SEI subgroup rescued the motility to at least half of the levels of the wildtype control (Figure 2.12C and 2.12D). As a co-receptor, BAK1 recognizes the Gly^{18} and Leu^{19} residues of the FLS2-bound flg22 (Sun et al., 2013). Since these residues are in the vicinity of Gln^{20} and Ile^{21}, we investigated how the perception of all the SEI variants impact FLS2-BAK1 interaction *in vitro*

(Smakowska-Luzan et al., 2018). We found that all of the niSEI peptides failed to trigger maximum heterocomplex formation, while the SEI peptides did not (Figure 2.5C). Thus, specific amino acid substitutions at position 18, 20 or 21 resolve AP by modulating the FLS2-BAK1 interaction. Notably, the change at position 20 offered a full bypass of AP in *Ps* (Figure 2.5B). Accordingly, *Ps* strains encoding a FliC variant carrying the flg22^{Q20D} epitope were able to colonize Arabidopsis significantly better than strains carrying Pa flg22 and inflicted more disease symptoms (Figure 2.5D-F). Since the changes at position 18, 20 or 21, did not affect the motility of *Pa* and *Ps* equally, we propose that the tunability of AP is dependent on the bacterial species considered.

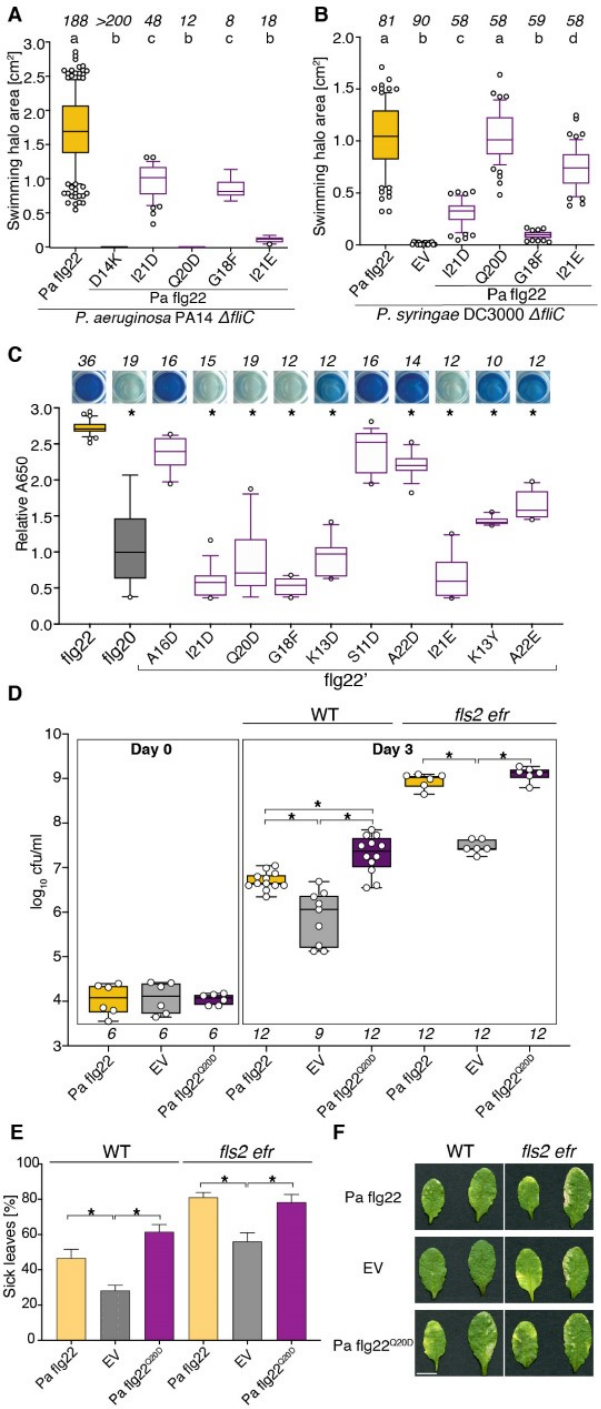


Figure 2.5. The tunability of antagonistic pleiotropy in the message segment of flg22 impacts the FLS2-BAK1 interaction and provides host colonization advantages. A and B- Quantification of swimming motility for *Pa* (A) and *Ps* (B) *ΔfliC* strains transformed either with an empty vector (EV) or with *fliC* variants coding Pa flg22, or the niSEI flg22 variants. Genotypes with the same letter are indistinguishable at >95% confidence (one-way ANOVA followed by Turkey HSD for all pairwise comparisons; $P < 0.05$). **C-** FLS2-BAK1 heterocomplex formation measured *in vitro* using an using an enzyme-linked immunosorbent assay. Represented is the relative absorbance (A650 nm) over 2 h. Representative pictures of wells are shown on top. For A,

B and C- the number of biologically independent observations (n) is indicated on top of each graph. **D**- Growth of *Ps ΔfliC* strains transformed either with an empty vector (EV) or with *fliC* variants coding Pa flg22, or flg22^{Q20D} on the plant genotypes indicated on top. The number of bacteria per area of leaf (cfu/ml) is plotted on a log10 scale for day 0 and day 3. Dots represent individual observations from two independent experiments. n = numbers of samples each including 8 biologically independent leaf discs is shown at the bottom. **E**- Quantification of leave disease symptoms using the bacterial strains and plant genotypes described above. For C, D and E- * indicates statistical significance (one-way ANOVA followed by a Dunnett's multiple comparison test; $P < 0.05$). **F-** Representative pictures of leaves taken 5 days after inoculation with the bacterial strains described above. Plant genotypes indicated on top. Scale bar: 1 cm.

2.2.5 Synthetic evidence for receptor antagonism as a strategy to dampen immune responses

Next, we obtained *in vitro* binding kinetics using GCI and determined that all of the niSEI peptides induced unstable FLS2-BAK1 complexes (Figures 2.6A and 2.13A; Data S3). We confirmed these results *in vivo* by using co-immunoprecipitation assays (Figure 2.6B). The FLS2/BAK1 interaction leads to the phosphorylation of the cytoplasmic kinase BOTRYTIS-INDUCED KINASE 1 (BIK1) for downstream signaling (Kadota et al., 2014; Lu et al., 2010; Wang et al., 2018). None of the niSEI peptides induced BIK1 phosphorylation (Figure 2.13B). Surprisingly, flg22^{A22E} triggered BIK1 phosphorylation without inducing FLS2-BAK1 interaction *in vivo* (Figures 2.6B and 2.13B). Thus, discrete polymorphic changes in flg22 can uncouple the responses of FLS2 in Arabidopsis. We propose that this uncoupling involves a regulatory mechanism that is yet to be identified. Consistent with this hypothesis, Colaianni and colleagues have demonstrated that a subset of commensal flg22 epitopes disconnect the signaling outputs of FLS2 (Colaianni et al., submitted).

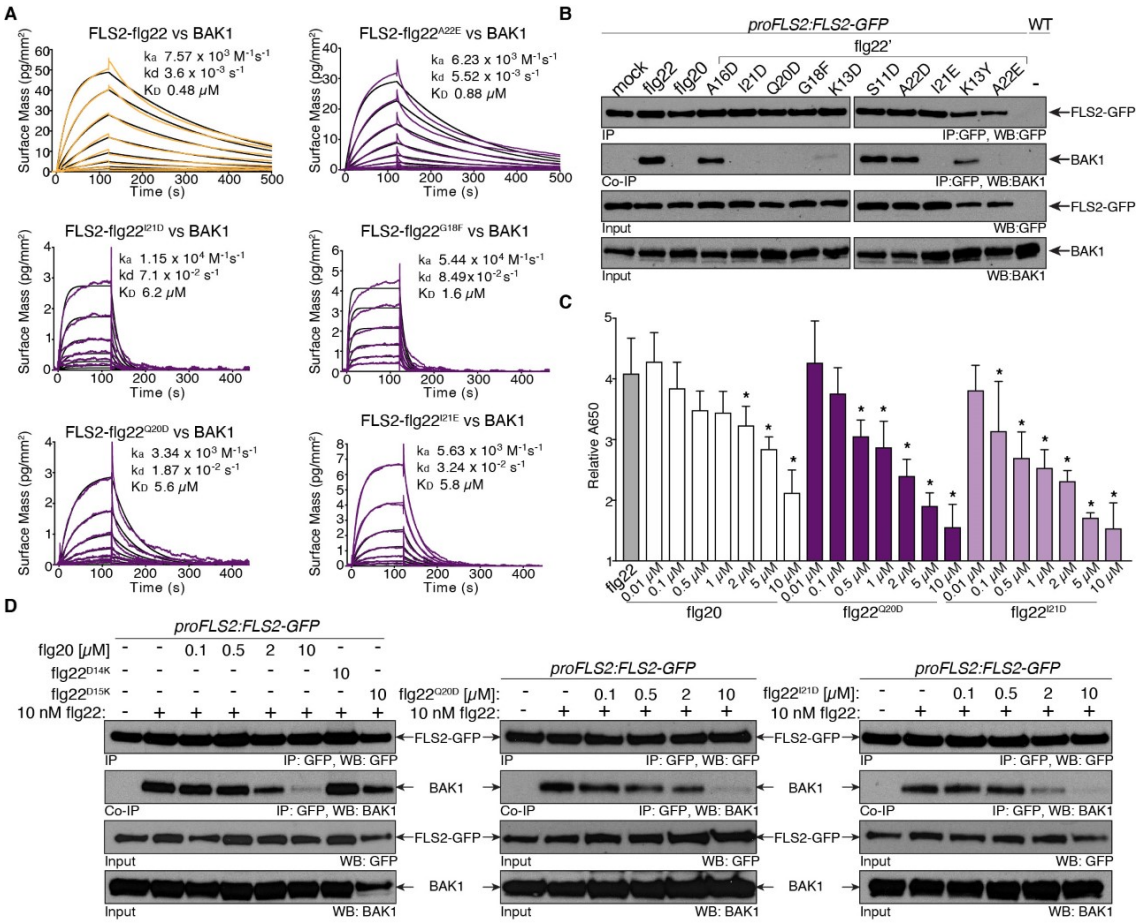


Figure 2.6. Competitive antagonism of receptor complex formation as a path to dampen immunity. A- Grating-coupled interferometry (GCI)-derived binding kinetics for the flg22-bound FLS2 and BAK1. Shown are sensorgrams with data in orange with the respective fits overlaid in black. The association (k_a) and dissociation (k_d) rate constants and the dissociation constant K_D are shown in the charts. **B-** Western blot analyses of FLS2–BAK1 co-immunoprecipitations (Co-IP/IP) in Arabidopsis transgenic seedlings (*proFLS2::FLS2-GFP)* either left untreated or treated for 10 min with different flg22 peptides. This experiment was repeated three times with similar results. **C-** FLS2-BAK1 heterocomplex formation measured *in vitro* using an enzyme-linked immunosorbent assay. Represented is the relative absorbance (Abs 650 nm) over 2 h obtained in the presence of Pa flg22 (10 nM) alone or at increasing concentrations of flg20, flg22^{Q20D} or flg22^{I21D}. The error bars indicate the standard deviation from at least two biological repetitions. * indicates statistical significance (one-way ANOVA followed by a Dunnett's multiple comparison test; $P < 0.05$). **D-** Western blot analyses of FLS2–BAK1 co-immunoprecipitations (Co-IP/IP) in Arabidopsis transgenic seedlings (*proFLS2::FLS2-GFP)* either left untreated, treated with Pa flg22 peptide (10 nM) alone, or co-treated with different flg22 peptides at the concentrations indicated on top. The identity of the flg22 peptides is indicated on the left. This experiment was repeated two times with similar results.

Next, we tested if the niSEI peptides could negate Pa flg22-induced activation of the immune marker gene *Cytochrome P450 71A12* (*CYP71A12*) (Millet et al., 2010). Using a

transcriptional reporter system, we determined that, except for flg22^{G18F}, all of the niSEI peptides antagonized the induction of *CYP71A12* by Pa flg22 (Figure 2.13C). In contrast, epitopes that interacted less with FLS2, were not able to block the induction of *CYP71A12* by Pa flg22 (Figure 2.13C). We then tested if flg22^{Q20D} and flg22^{I21D} antagonized the formation of the FLS2-BAK1 heterocomplex induced by Pa flg22 *in vitro* (Figure 2.6C). We determined that both peptides readily inhibited this interaction at concentrations fifty times higher than Pa flg22. In contrast, Pa flg20 required a 200-fold molar excess, to achieve the same level of antagonism. Finally, we demonstrated that flg22^{Q20D} and flg22^{I21D}, but not flg22^{D14K} or flg22^{D15K}, antagonized the flg22-induced FLS2-BAK1 interaction *in vivo* (Figure 2.6D). Thus, we synthetically evolved antagonistic flg22 epitopes that inhibit the activation of downstream FLS2 responses with no apparent effect on motility.

2.2.6 Acquisition of antagonist peptide function as a natural strategy to poison FLS2 function

In a next step, we looked for the occurrence of the substitutions identified by our synthetic approaches in natural lineages of *Pseudomonas*. We found that Asp^{15}, but not Asp^{14}, occurred as Gly (flg22^{D15G}) or Asn (flg22^{D15N}) in *Pseudomonas fluorescens* and *Ps* pathovars that associate with plants (Winsor et al., 2016). Since both mutations reduced the motility of *Pa* (Figure 2.10C), we hypothesized that variations of the Asp^{15} residue would perhaps be accompanied by concomitant changes that compensate the presumed motility defect. To test this, we used mutual information theory (MIT) to analyze the co-occurrence of polymorphisms in FliC protein sequences extracted from the general Pseudomonas Genome Database (Figure 2.6A; Table S3) (Gouveia-Oliveira and Pedersen, 2007; Winsor et al., 2016). We found that co-variation between positions 15 and both 17 and 18 in the flg22 segment distributed in the top 0.5%

of MIT scores (Figures 2.7A). We retrieved two representative epitopes with these co-variations and focused on them for experimental validation. These epitopes had additional polymorphisms in the 'address' and 'message' segments and were named flg22^{GDG} (Gly^{15}/Asp^{18}/Gly^{20}) and flg22^{NSS} (Asn^{15}/Ser^{17}/Ser^{22}). Then, we set out to understand the functional impact of this polymorphism on immunogenicity (Figure 2.7B). As controls, we generated chimeric peptides (flg22^{chGDG} and flg22^{chNSS}) in which the message segments were fused to the address of Pa flg22 and subsequently tested their ability to induce ROS bursts and FLS2- BAK1 interaction *in vitro* (Figure 2.7B and 2.14A). While flg22^{NSS} and flg22^{chNSS} induced normal FLS2/BAK1 interactions and wildtype ROS bursts, flg22^{GDG} and flg22^{chGDG} were unable to do so (Figure 2.7B and 2.14A). The results obtained for flg22^{NSS}/flg22^{chNSS} imply that co-variation between position 15 and 17 is perhaps not sufficient to avoid the induction of FLS2 responses in Arabidopsis.

To understand why flg22^{chGDG} is non-immunogenic, we synthesized Pa flg22 peptides in which the substitutions were introduced separately (flg22$^{D15G,\ G18D}$ and ^{Q20G}) or in pairwise combinations (flg22$^{D15G/G18D,}$ flg22 $^{D15G/Q20G}$ and flg22$^{G18D/Q20G}$), and tested ROS responses (Figure 2.7C). The ROS bursts mediated by flg22^{D15G} were indistinguishable from that of Pa flg22. These results are supported by the slight reduction in FLS2 interaction observed on our peptide array for this variant (Figure 2.2D). Remarkably, flg22^{G18D} and flg22^{Q20G} induced lower ROS burst responses than Pa flg22, and these responses were further decreased when the mutations occurred together (flg22$^{G18D/Q20G}$) or jointly with flg22^{D15G} (Figure 2.7C). Thus, we propose that positive epistatic interactions between these residues occur to greatly reduce epitope immunogenicity. Next, we reasoned that the introduction of a negative charge at position 18 should rescue the slight reduction in FLS2 interaction for flg22^{D15G}. Second, we expected that the introduction of this negative charge together with the change at position 20 would result in a

molecular phenocopy of the niSEI epitopes flg22^{G18F} and flg22^{Q20D} (Figures 2.5C, 2.6C and 2.6D), thereby allowing flg22^{GDG} to antagonize the Pa flg22-induced FLS2-BAK1 interaction. As expected, flg22^{GDG} and its corresponding chimera failed to trigger FLS2-BAK1 heterocomplex formation, and antagonized the activities of Pa flg22, both *in vitro* and *in vivo* (Figures 2.7D, 2.7E, 2.14A and 2.14B). We propose that positive epistasis not only channels the positive pressure acting to mediate evasion from FLS2 detection in Arabidopsis, but also dampens plant immune system function by poisoning the activity of FLS2-BAK1 heterocomplexes with antagonist epitopes.

To establish the effects of positive epistasis on flagellin function, we assessed the impact of the flg22^{GDG} polymorphism on the motility of *Pa* and *Ps* either individually or collectively (Figure 2.7F, 2.14C and 2.14D). While FliC variants with flg22^{GDG} epitopes failed to rescue either *Pa* and *Ps ΔfliC,* the chimera restored 12.5 % of motility of *Ps* but not in *Pa* (Figures 2.7F and 2.14D). These results reinforce the notion that the tunability of motility is species dependent (Figure 2.5A and 2.5B). When introduced separately, each of the mutations restored 18.5 % of the motility of *Pa ΔfliC* (Figure 2.14C). Thus, the residues 15, 18 and 20 can interact negatively to control *Pa* motility. Control complementation assays performed with flg22^{NSS} derivatives yielded nearly similar results (Figures 2.7F and 2.14D), and further indicated that *Pseudomonas* motility is very sensitive to residue co-variations in flg22. The natural polymorphism at position 20 (flg22^{Q20G}), as for the niSEI peptide flg22^{Q20D} (Figure 2.5B), offered the most balanced resolution of AP (Figures 2.7C and 2.14B). Altogether, our results indicate that positive epistatic interactions mediating evasion from immune system detection lead to negative effects on flagellin function. We propose that flg22^{GDG}- like polymorphisms could eventually persist on

Arabidopsis because the benefits of reducing immune system detection via antagonist activity outweighs the negative effects on motility function.

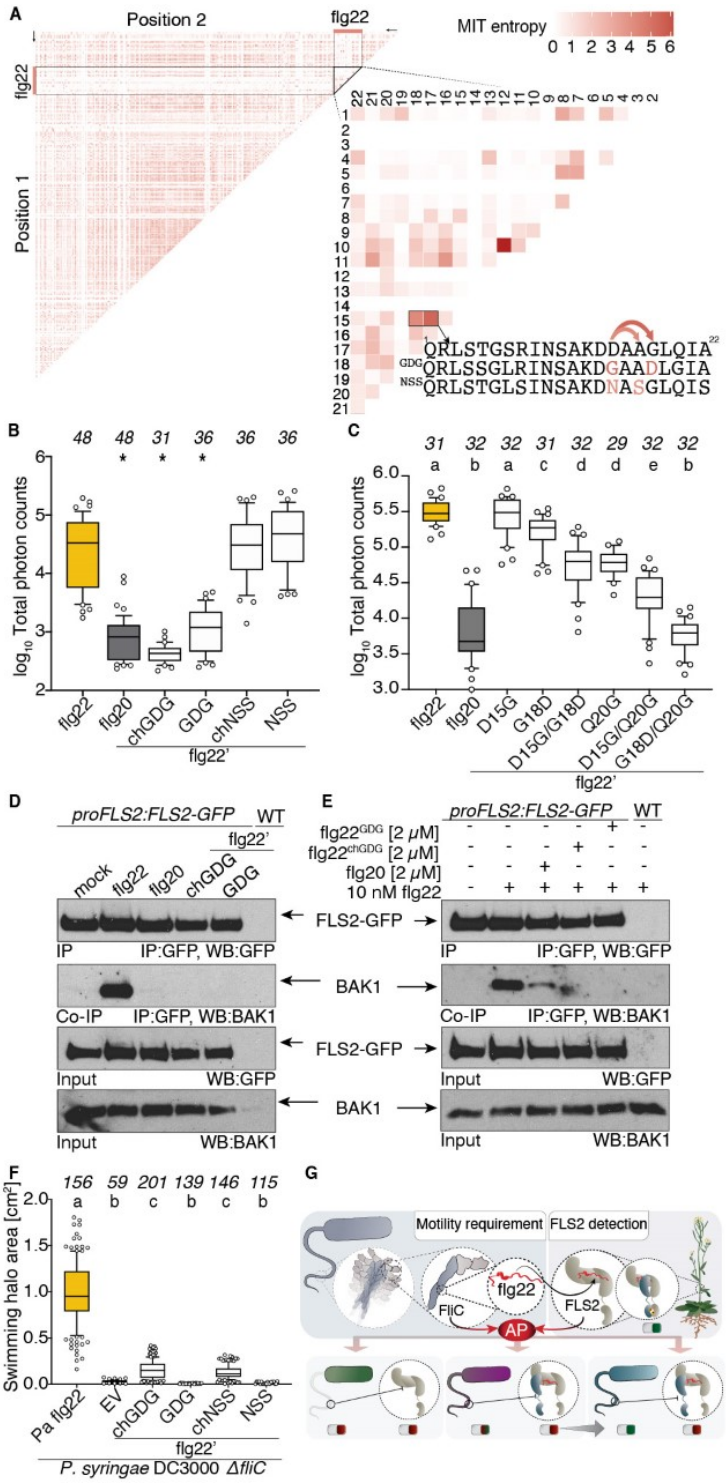


Figure 2.7. Natural occurrence of antagonist epitope function in *Pseudomonads*. A- Mutual information theory (MIT) heat map of FliC protein sequences from the Pseudomonas Genome

Database. The flg22 segment is magnified on the right. The colour scale bar on top shows the MIT entropy score values between positions. The arrows on the flg22 sequence at the bottom indicate that Asp^{15} co-varies with Ala^{17} ($flg22^{NSS}$) or Gly^{18} ($flg22^{GDG}$). **B and C-** flg22-induced oxidative bursts represented as total photon counts over 40 mins. Pa flg22, $flg22^{GDG}$, $flg22^{NSS}$ and the respective chimeras $flg22^{chGDG}$/$flg22^{chNSS}$ (**B**) along with peptides that recapitulate different combinations of single and pairwise polymorphism in $flg22^{chGDG}$ variant (**C**) were used at a concentration of 100 nM. For B- * indicates statistical significance (one-way ANOVA followed by a Dunnett's multiple comparison test; $P < 0.05$). **D-** Western blot analyses of FLS2– BAK1 co-immunoprecipitations (Co-IP/IP) in Arabidopsis transgenic seedlings (*proFLS2::FLS2-GFP)* either untreated (mock) or treated for 10 min with the flg22 peptides indicated on top. This experiment was repeated two times with similar results. **E-** same as (D) except that the seedlings were treated with 10 nM Pa flg22 alone; or co-treated with 10 nM Pa flg22 and 2 μM of $flg22^{GDG}$/$flg22^{chGDG}$. This experiment was repeated two times with similar results. **F-** Quantification of swimming motility of *P. syringae ΔfliC* strains transformed either with an empty vector (EV) or with *fliC* variants coding either Pa flg22, $flg22^{GDG}$, $flg22^{NSS}$ or the respective chimeras $flg22^{chGDG}$ /$flg22^{chNSS}$. C and F- Peptides and genotypes with the same letter are indistinguishable at >95% confidence (one-way ANOVA followed by Turkey HSD for all pairwise comparisons; $P < 0.05$). The number of biologically independent observations (n) is indicated on top of each graph. **G-** Model depicting the role of antagonistic pleiotropy (AP) on the functionality of FliC. AP maintains allelic diversity in flg22 by balancing the negative pressure to maintain FliC function against the positive pressure to evade FLS2 detection. Resolution of AP lead to (1) epitopes with reduced agonist function and dramatic reduction in motility (green bacteria), or to (2) epitopes with antagonist function and moderate reduction of motility (purple bacteria). Natural resolution of AP occurs through similar mechanisms (blue bacteria).

In support of this, we found that the $flg22^{GDG}$ allele occurred in the genomes of *Pseudomonas* operational taxonomical units (*Ps* OTUs) that stably associate with Arabidopsis in the wild (>1500 genomes; Karasov et al., 2018). Because the $flg22^{GDG}$ allele occurs at very low frequency in these genomes (<0.1%; Table S4; Data S1), we propose that this allele is nonadaptive for Arabidopsis colonization. In contrast, we found that >85% of these genomes encoded two distinct flg22 epitopes reminiscent of the antagonist niSEI peptides, with changes present at position 20 and 21 (Table S4). In addition to these substitutions, a bit less than half of the OTUs genomes analyzed with the MIT algorithm showed co- variation between position 15 and 18 in flg22 (Figures 2.14E and 2.14F; Table S5). Thus, we used co-variation as a criterion to classify these peptides; the flg22 peptide with co-variation between Asn^{15} and Val^{18} accompanied by substitutions at positions 20 and 21, was named Var I ($flg22^{V.I}$) (Figure 2.14E);

the other peptide with only substitutions at positions 20 and 21 was named Var II ($flg22^{V.II}$) (Figure 2.14E). In ROS burst assays both epitopes proved non-immunogenic (Figure 2.14G). This is consistent with previous results showing that Var I do not induce SGI (Vetter et al., 2016). Importantly, the change into Asn^{15} in the context of the Pa flg22 sequence did not affect ROS bursts responses (Figure 2.14G). Thus, substitutions at position 15 do not necessarily impact FLS2 responses. Rather, we propose that the loss-of-immunogenic function in Var II are mediated by the changes at position 20 and 21. Based on the data obtained for the niSEI and $flg22^{GDG}$ epitopes, we hypothesized that Var I and Var II would both act as FLS2 antagonists. Expectedly, Var I and Var II reduced Pa flg22-induced FLS2-BAK1 interaction *in vivo,* albeit with different strengths (Fig. 2.14H). When compared to Var II, Var I more strongly inhibited Pa flg22-induced FLS2-BAK1 complex formation (Fig. 2.14H). Our results are supported by the antagonistic effects of Var I and Var II on Pa flg22 induction of *CYP71A12* (Colaianni et al., submitted). We propose that the co-variation between position 15 and 18 in Var I, interacts positively with the changes at position 20 and 21 to increase antagonist function. Because single substitutions to Asn^{15} or to Val^{18} in Pa flg22 reduce the motility of *Pa* by approximately 70% (Fig. 2.14I), it is likely that the polymorphism of Var I and Var II lead to reduction in motility. Alternatively, this motility reduction is either minimally adaptive and/or compensated by co-evolving sites outside the epitope segment (Figure 2.14F). We hypothesize that the reduction of FLS2 function via antagonist activity has taken priority over the maintenance of motility in the *Ps* OTUs, thereby contributing to the dominance of this lineage on Arabidopsis.

Next, we sought for genomic evidences supporting that the design of epitopes with FLS2-BAK1 antagonist activities could provide widespread advantages for Arabidopsis colonization. For this, we analyzed the polymorphism of *FLS2* and *BAK1* in genomes of >1,000 natural

Arabidopsis accessions (Alonso-Blanco et al., 2016). Our analysis revealed that >100 non-synonymous changes can occur in $FLS2^{ECD}$ (Table S6). Of these, 41 substitutions involved amino acids located in leucine-rich repeats (LRRs) that contribute directly to interaction with flg22 (LRR2-16) (Sun et al., 2013). Arabidopsis possibly uses this variation to refine FLS2 selectivity for distinct types of epitopes. Consistent with this, FLS2 exhibit distinct epitope selectivity in land pants (Fürst et al., 2020; Robatzek et al., 2007; Takai et al., 2008; Trdá et al., 2014; Wei et al., 2020). In contrast to $FLS2^{ECD}$, polymorphism in $BAK1^{ECD}$ is nearly non-existent, and only one non-synonymous change occurs in the extracellular juxtamembrane domain (Table S6). BAK1 physically interacts with over 30 receptor kinases to regulate other immune and developmental pathways (Boutrot and Zipfel, 2017; Hohmann et al., 2017; Smakowska-Luzan et al., 2018). Since single amino acid changes in the $BAK1^{ECD}$ can disrupt the functioning of important signaling pathways (Jaillais et al., 2011), we attribute the lack of polymorphism to this 'hub' function. Based on this, we posit that dominant isolates from the *Ps* OTUs have evolved to make FLS2 polymorphism irrelevant by exploiting the constrained 'hub' function of BAK1 in Arabidopsis. This is consistent with our demonstration that >85% of these Arabidopsis isolates encode epitopes that antagonize the activities of Pa flg22 on the FLS2-BAK1 heterocomplex. The results of Colaianni and colleagues further indicate that the acquisition of flg22 epitopes with antagonist function is not limited to *Pseudomonads* but also occurs in alpha- and betaproteobacteria that are found in Arabidopsis microbiomes (Colaianni et al., submitted).

Section 2.3: Discussion

Evolutionary mechanisms that help explain how natural bacterial communities maintain core microbial functions while circumventing immunity are scarce (Teixeira et al., 2019), mainly because direct molecular evidences are challenging to obtain experimentally. Here, we

demonstrate that AP refines the relationship between an epitope and an immune sensor and, in turn, show that tradeoffs between bacterial motility and host recognition have been molecularly balanced throughout evolution.

The motility function encoded in flg22 is selectively monitored by FLS2. Our results indicate that >300 single alterations of flg22 have a drastic impact on *Pa* motility. Collectively, our data demonstrate that flg22 has, to a greater extent than anticipated, a coding function that massively impinges on *Pseudomonads* motility whereby amino acid changes targeting less-conserved residues also caused drastic motility defects. Conversely, mutations of some invariant amino acids had no obvious effects on motility. Thus, the evolutionary fixation of residues in the epitope is unlikely to be driven by selective mechanisms aimed only at maintaining bacterial motility. Our systematic analysis of flg22-FLS2 interactions, allowed us to determine that in Arabidopsis, FLS2 did not evolve to monitor all the epitope positions that are necessary for motility. We propose that natural Arabidopsis accessions have evolved this capacity in their flagellin sensor to fine-tune and prioritize their defense responses against different bacteria. Consistent with this, natural flg22 epitopes induce diverse immune responses (Colaianni *et al.*, submitted).

FLS2 resiliently interact with mutated flagellin epitopes. The directed evolution of >1000 epitopes has produced novel insights that expand our understanding of the FLS2/flg22 interaction in important ways. While our data are mostly aligned with the crystallographic data (Sun et al., 2013), we have found that, except for mutations targeting the $Asp^{14/15}$ doublet, Gln^{20}, and Ile^{21}, all other mutations targeting residues with side chains that physically interact with $FLS2^{ECD}$ in the crystal structure had no major impact on the interaction. Most notably, this robust interaction is only modulated by the introduction of residues with opposed electrostatic charges in

the C-terminal moiety of the epitope. Remarkably, these electrostatic effects were magnified on 15-mer derivatives of the epitope that are universally recognized as immunogenic by FLS2 orthologs in flowering plants (Bauer et al., 2001; Felix et al., 1999; Robatzek et al., 2007). Overall, we propose that in Arabidopsis the robust interactions of FLS2 with mutated flagellin epitopes establishes AP and contributes to the maintenance of flagellin epitope variation in bacterial populations (Colaianni et al., submitted).

Antagonistic pleiotropy reduces the potential adaptive value of FliC. An obvious regulatory solution to the problem of AP would be to simply eliminate *fliC*. However, the genetic elimination of *fliC* in various plant associated bacteria significantly reduces their virulence (Cesbron et al., 2006; Chesnokova et al., 1997). Bacteria employ other mechanisms to provide resolution of AP. One tactic deployed by several *Pseudomonads* involves degradation of excess flagellin monomers by secretion of proteases in the extracellular space (Bardoel et al., 2011; Pel et al., 2014). Another tactic relies on flagellin glycosylation (Taguchi et al., 2010). In some plant-specific bacteria, the addition of unique glycan structures prevents the appropriate release of epitopes, thereby leading to non-immunogenic flagellins (Buscaill et al., 2019). The most direct resolution of AP could be provided by the positive selection pressure driving discrete levels of polymorphism in the flg22 sequence (Felix et al., 1999; McCann et al., 2012; Mott et al., 2016; Sun et al., 2006). Yet, only a few pathogenic species seem to utilize this strategy because reductions in motility are suboptimal for host adaptation. In comparison, a relatively high-level of polymorphism is maintained in commensal communities (Colaianni et al., submitted). This may indicate that these communities are less dependent on flagella-mediated motility and/or maintain multiple *fliC* copies in their genomes to resolve AP (Colaianni et al., submitted). Alternatively, immune suppression strategies such as these mediated by the type III

secretion system and its effector proteins could be used to alleviate pressure on *fliC*. In this strategy, however, resolution of AP would only be transferred to other processes since bacterial effectors are also robustly detected by plants Nod-Like Receptors (Laflamme et al.,2020).

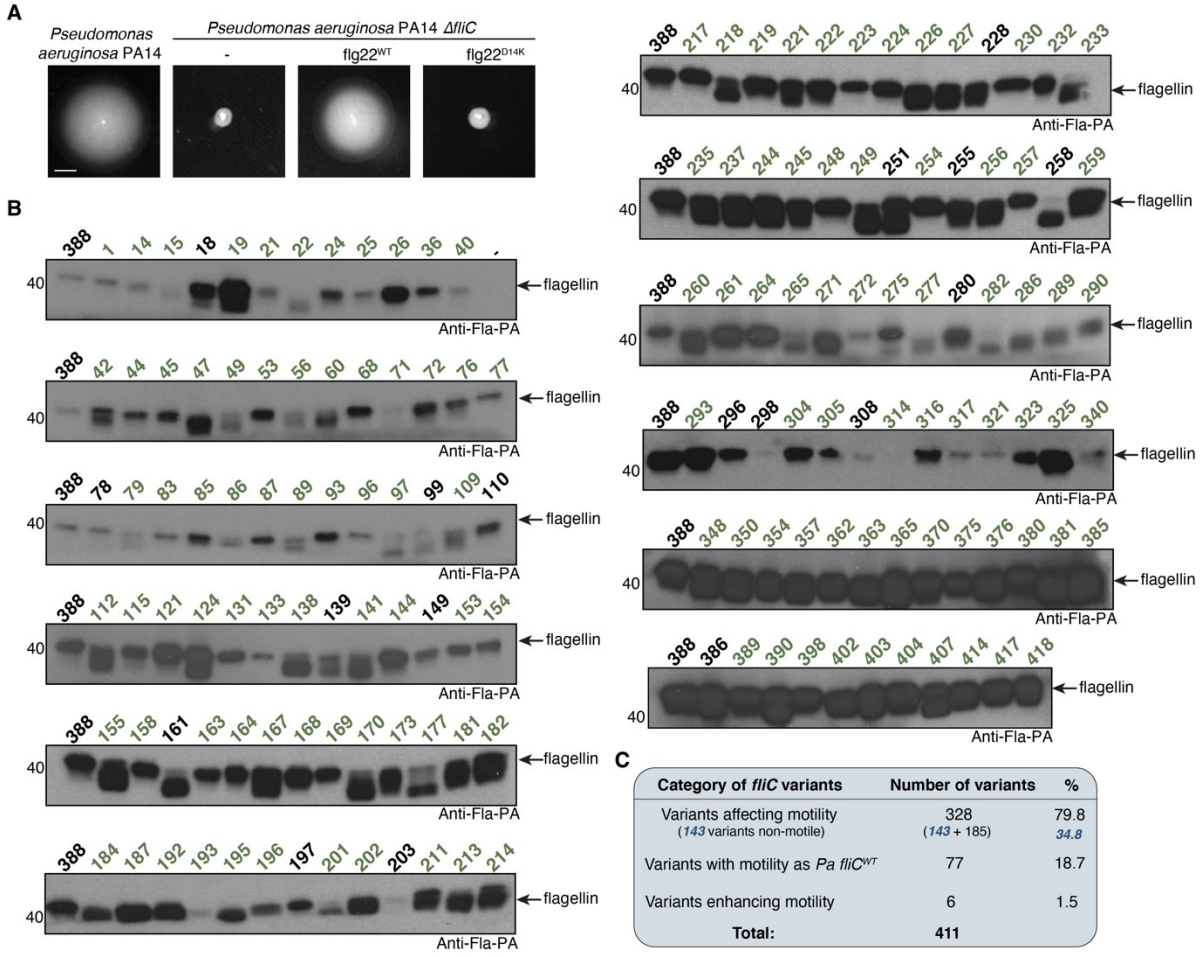
Mutational scenarios for the acquisition of antagonist epitopes in *Pseudomonads*. We have identified mutational tracks in a flagellin epitope that drive plant immune system evasion and poisoning by antagonist production (Figure 2.7G). We found signatures of these tracks to be layered atop each other in naturally occurring bacteria that associate with Arabidopsis in the wild. While we cannot precisely determine how these mutational signatures emerged, we propose that they were acquired through continuous series of mutagenic events. First, we envision that natural selection directed evasion from immune system detection and as a consequence stalled evolution by establishing AP. A second series of selection events only parsimoniously resolved AP but led to epitopes that reduce immune detection via antagonistic function. This newly acquired function persisted because it provides greater benefits for bacterial colonization than a mere restoration of motility (Figure 2.7G). Amongst others, a likely scenario could have included some of the following events; (1) natural selection has favored specific substitutions at position 15 in a background of mutations targeting other residues in the message segment (e.g position 20); hence, resulting in bacteria retaining adaptive levels of motility while carrying epitopes with suboptimal reductions in immunogenicity, (2) the strong negative selection on BAK1 in Arabidopsis channeled frequent parallel co-variation between position 15 and 18, and beneficial epistatic interactions between these residues and the residue at position 20 allowed a full reduction in epitope immunogenicity,

(3) together these substitutions are also beneficial because they generate weak receptor antagonists and as such prevent an over titration of BAK1 function on other MAMP responses;

hence, allowing harmonious microbiome assembly by avoiding 'over-colonization' of the host by other competing species (Lee and Belkhadir, 2020; Smakowska et al., 2016), (4) despite drastic effects on motility, this combination of substitutions was positively selected to allow bacteria carrying such epitopes to persist on Arabidopsis. Since the polymorphism at position 20, can offer a balanced resolution of AP in *Ps*, substitutions that occurred first at position 20 (or in other key residues of the 'message' segment) and were followed by co-variations at position 15 and 18 are also conceivable. In support of this second scenario, position 20 is the most polymorphic in flg22 epitopes encoded by *Ps* OTUs strains.

All in all, we provide a rationale to better understand the molecular of evolution of other MAMP-receptor pairs in plants but also in lower and higher metazoans. For instance, our study could be used as a template to dissect how bacterial protein synthesis mediated by the Elongation factor Tu (EF-Tu) has evolved in the face of recognition by the Brassicaceae-specific EF-Tu receptors (EFR) (Boutrot and Zipfel, 2017). Understanding how sequence variation in EF-Tu epitopes has been influenced by AP to avoid immune detection while maximizing ribosome function during the elongation phase of translation would be of great interest. As a whole, the concept explored has the potential to reframe our evolutionary understanding of non-self perception.

Section 2.4: Supplemental figures



Category of *fliC* variants	Number of variants	%
Variants affecting motility (*143* variants non-motile)	328 (*143* + 185)	79.8 *34.8*
Variants with motility as *Pa fliC*^{WT}	77	18.7
Variants enhancing motility	6	1.5
Total:	**411**	

Figure 2.8. Flagellin is expressed in *Pseudomonas* strains with reduced motility, Related to Figure 2.1. A- Pictures showing the motility phenotype of *Pseudomonas aeruginosa*, Δ*fliC* and Δ*fliC* strains complemented with *fliC* variants coding the flg22^{WT} (ID #388) or flg22^{D14K} epitopes. Scale bar: 0.5 cm. **B-** Western blot analysis of flagellin expression in *P. aeruginosa fliC* allelic clones with either reduced (ID number marked in black) or no motility (ID number marked in green). Anti-flagellin antibodies were used for analysis (Anti-Fla-PA). **C-** Table related to Figure 2.1B. Percentages of *fliC* transformants with either a significantly reduced, enhanced or unaffected motility when compared *fliC* wild-type (WT).

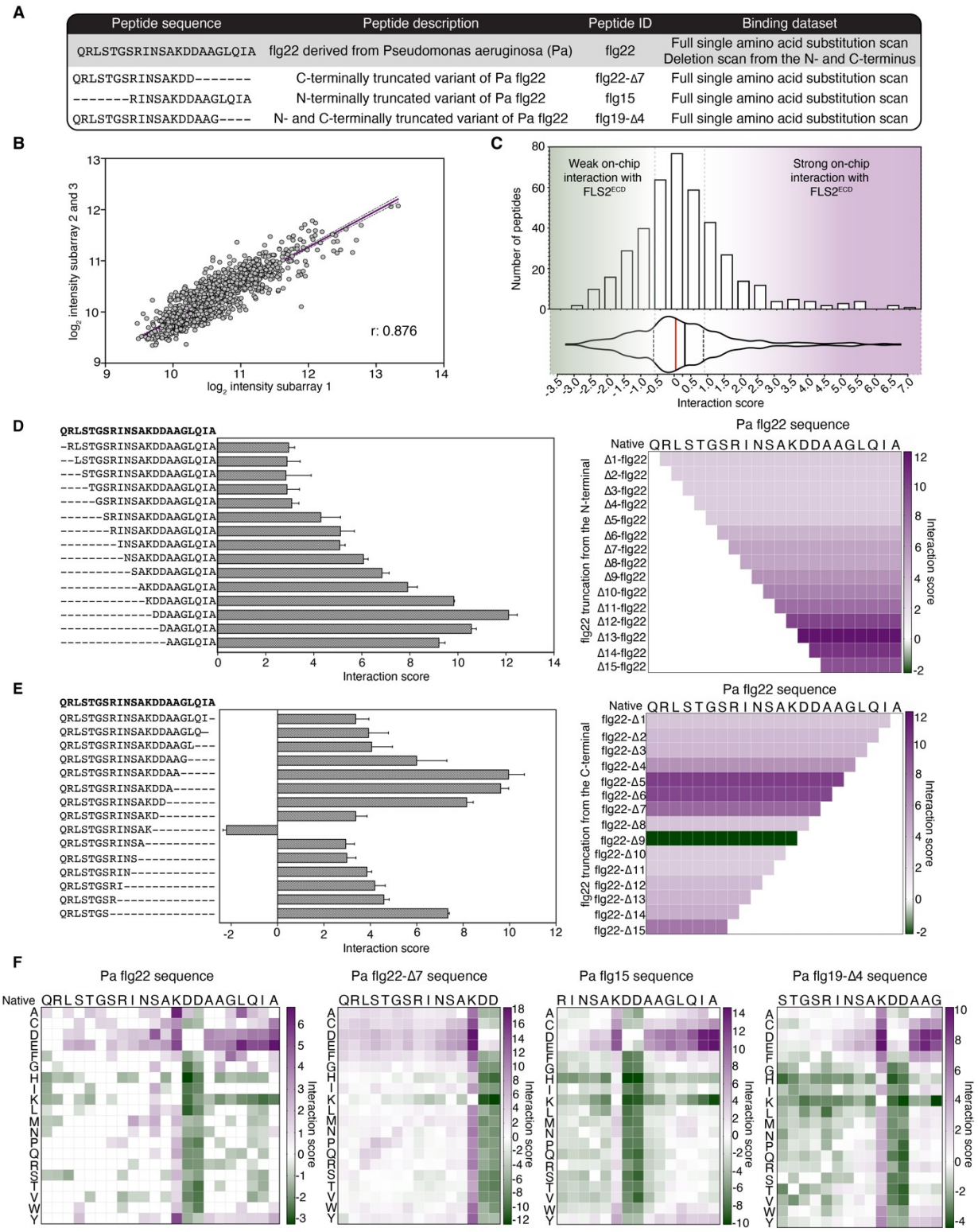


Figure 2.9. Analysis of the flg22/ FLS2^{ECD} interaction results for the mutational scans targeting full-length and truncated epitopes, Related to Figure 2.2. A- Summary table of binding datasets obtained for FLS2^{ECD} interaction with derivatives of four flg22 variants. **B-** Scatter plots showing the correlation of fluorescent intensities between replicates distributed over

three sub-arrays of the peptide chip. Correlation between the full-length Pa flg22 mutational scans are presented. The purple line is the linear regression that best describes the dataset. The dashed lines comprise 95% confidence intervals for the fit (r: Spearman's correlation coefficient). **C-** Histogram with a corresponding violin plot for the interaction scores summarized in Figure 2.2A. The first and third quartiles are marked with dashed lines on the violin plot. The median and mean values of the data are indicated by red and black lines, respectively. **D and E-** Bar plot (left panel) and heatmap (right panel) showing the impact of truncations of flg22's extremities on the interaction with $FLS2^{ECD}$. **F-** Position specific scoring matrices (PSSMs) representing the impact of each amino acid substitution on the interaction of $FLS2^{ECD}$ with full length Pa flg22 or with epitope variants truncated at either or both of the N- and C-terminal moieties. Substitutions generating peptide variants with interaction scores below or above the non-mutated reference flg22 peptide are shown in green or purple, respectively.

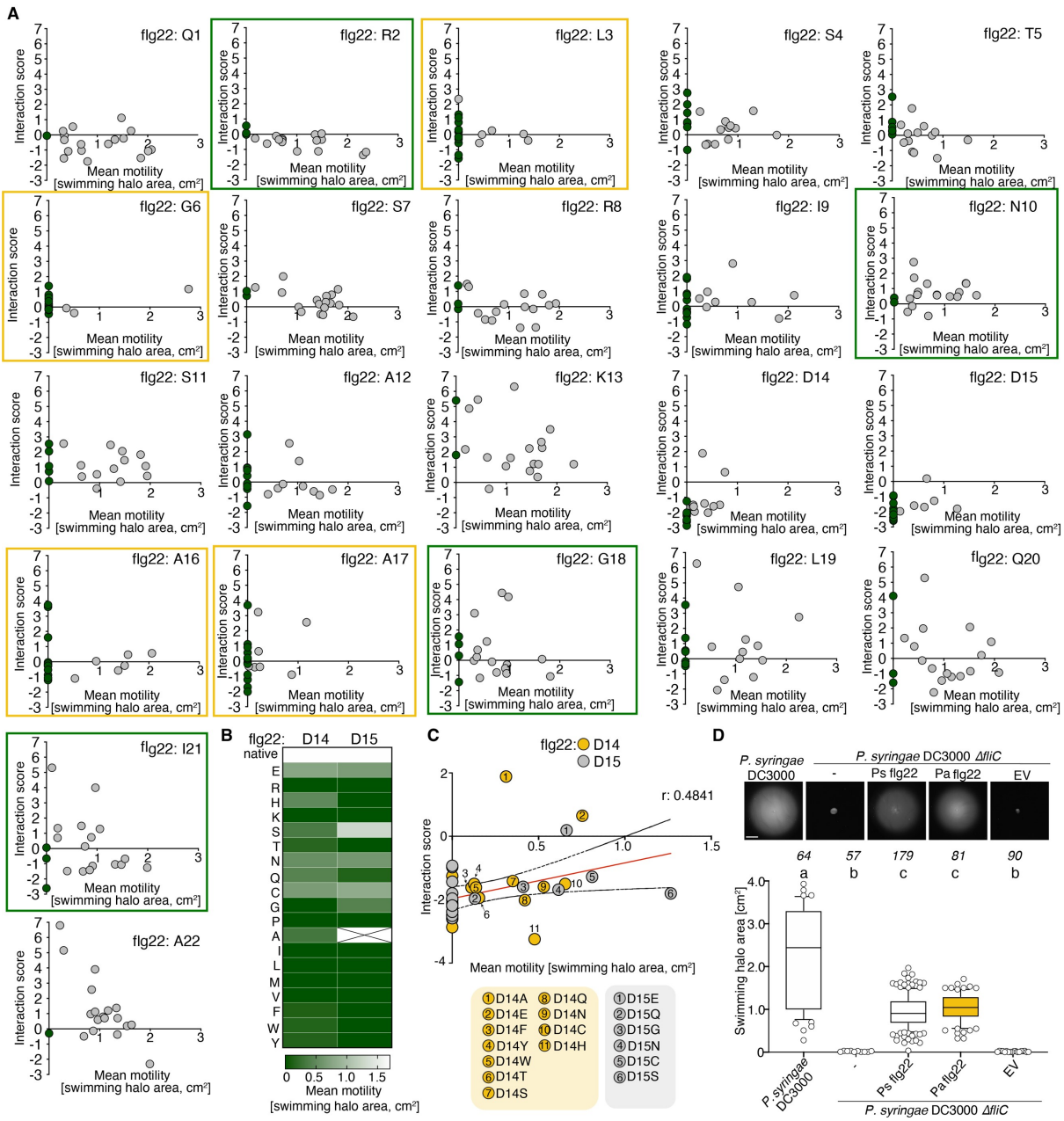


Figure 2.10. Mutations of the Asp$^{14/15}$ doublet have the largest impact on bacterial motility and interaction with FLS2, Related to Figure 2.3. A- Scatter plots presenting the impact of replacements at each amino acid on the motility of *Pa* transformants (x-axis) and the interaction with FLS2^{ECD} (yaxis). The green dots indicate mutations that cause a complete loss of motility. Invariant or highly conserved amino acids for which mutations have major impacts on motility and only minor effects on FLS2 interaction are surrounded by an orange square. Invariant or highly conserved amino acids for which mutations have minor impacts on motility and FLS2 interaction are surrounded by a green square. **B-** Heatmap showing the impact on *Pa* motility for each amino acid replacement targeting the Asp$^{14/15}$ doublet. The mutation corresponding to the white box (D15A) was not tested. **C-** Correlation between the motility of *Pa* transformants (x-axis) and the FLS2^{ECD} interaction scores (y-axis) for mutations targeting Asp^{14} (orange dots) or Asp^{15} (grey dots). The numbers in the dots correspond to the amino acid changes presented in the table below

the scatter plot. The red line is a linear regression that best describes the dataset and the dashed lines comprise 95% confidence intervals for the fit (r: Spearman's correlation coefficient). **D -** *P. syringae* strains coding chimeric *fliC* variants with *P. aeruginosa* flg22 epitopes are motile. **Top panel:** Pictures showing the motility phenotype of *Ps* DC3000 WT, *Ps* Δ*fliC* and strains transformed with an empty vector (EV) or with a vector coding *fliC* variants with the Ps flg22 and Pa flg22 epitopes. Scale bar: 0.5 cm. **Bottom panel:** Quantification of swimming motility. The box plot represents the first and third quartiles, centered by the median. Whiskers include the 10^{th}-90^{th} percentile of the data points from at least three independent biological experiments. Statistical significance was assessed using one-way ANOVA followed by Tukey's HSD for all pairwise comparisons ($P < 0.05$). Genotypes with the same letter are indistinguishable at >95% confidence. The number of biologically independent observations (n) is indicated on top of the graph.

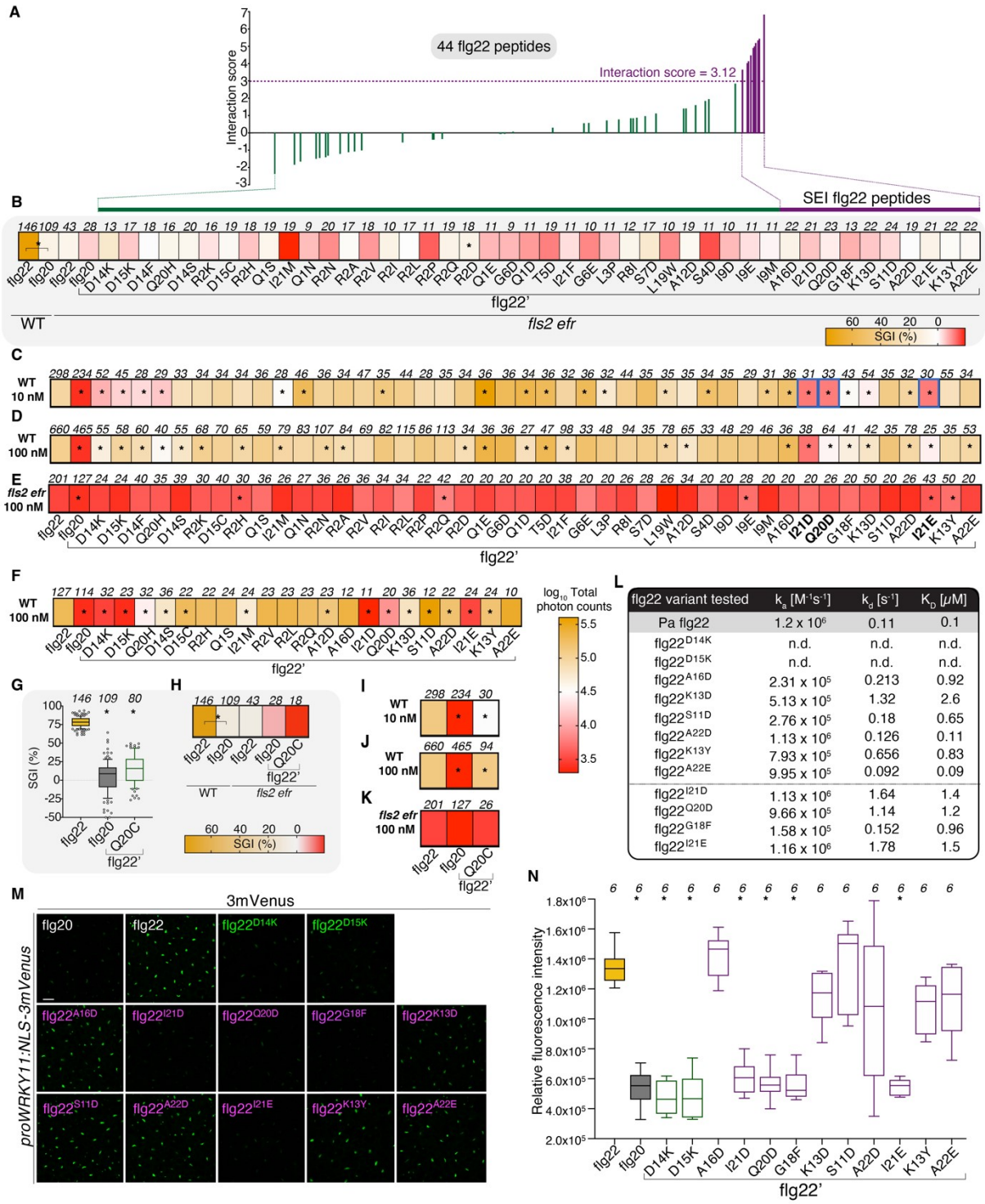


Figure 2.11. Analysis and verification of flg22 peptide immunogenicity *in planta*, Related to Figure 2.4. A- The flg22 peptides selected for functional validation *in planta* uniformly cover the FLS2 interaction score distribution. 10 out of 44 peptides show a significantly enhanced interaction (SEI) with $FLS2^{ECD}$ and are marked in purple. The remaining peptides are marked in green. **B-** flg22-induced seedling growth inhibition (SGI) in *fls2 efr* mutant plants. Shown is a heat map of SGI responses. The color scale bar at the bottom shows the SGI percentage values. Pa flg22-

induced SGI in WT plants is marked in orange and was used as reference to build the heatmap. **C, D, E and F -** flg22-induced reactive oxygen species (ROS) burst in WT (C, D and F) and *fls2 efr* (E) mutant plants. Shown is a heat map of ROS responses. The color scale bar on the right next to panel F shows the Log10 transformed total photon counts. For (C), the ROS burst measured for each peptide were compared to flg20's photon count using a Monte-Carlo approach. flg20's datapoints were randomly downsampled to an n equal to the peptide it was being compared to. Then 95% confidence intervals were computed and the number of times they overlapped was recorded for 10,000 downsamples. If the 95% confidence interval overlapped on less than 5% of the downsamples, the pixels corresponding to the peptide's response were surrounded with a blue box. These peptides ID is highlighted in bold and at the bottom of the chart. For (F), the flg22-induced ROS bursts were measured using a different luminol reagent. For B-F, the peptides are ordered from left to right by increasing interaction score with FLS2. **G and H-** flg22-induced SGI in WT (G) and *fls2 efr* mutant plants (H) for $flg22^{Q20C}$. **I, J and K-** flg22induced reactive oxygen species (ROS) burst in WT (I, J) and *fls2 efr* (K) mutant plants for $flg22^{Q20C}$. For B-K, the number of biologically independent samples (n) are indicated on top. The identity of the flg22 peptide tested is indicated at the bottom. Statistical significance was assessed using a linear mixed model to compare the response of each flg22 variant to the Pa flg22 control response using an unpaired two-sided *t*-test followed by multiple testing correction using the Holm method (*$P < 0.05$). **L-** Table summarizing the Grating-coupled interferometry (GCI)-derived binding kinetics for the interaction between $FLS2^{ECD}$ and the flg22 variants indicated on the left. The corresponding association (k_a) and dissociation (k_d) rate constants and the dissociation constant K_D are indicated. The sensorgrams with raw data and the respective fits are presented in (Data S4). **M-** Expression patterns of the defense marker gene *WRKY11* in response to flg22 epitope treatments in epidermal cells of cotyledons. Shown are images from single confocal sections. The epitopes are indicated on top of each panel. Scale bar for all pictures is 50 µm. **N-** Quantitative analysis of mVENUS signal intensities after treatment with the flg22 peptides indicated at the bottom. The number of biologically independent cotyledons (n) is indicated on top. Statistical significance was assessed using a one-way ANOVA followed by a Dunnett's multiple comparison test (*$P < 0.05$). For G and N, the box plot represents the first and third quartiles, centered by the median. Whiskers include the 10^{th}-90^{th} percentile of the data points. The boxplots are marked with the following color code: wildtype flg22 (in orange), flg20 (grey), SEI variants (purple), all others (green).

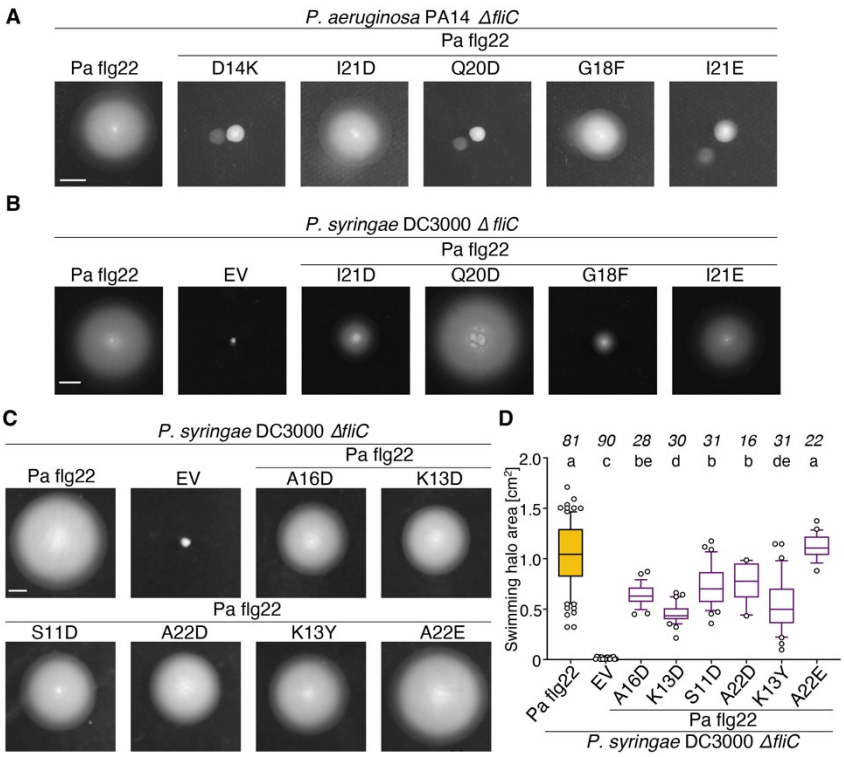


Figure 2.12. Fine-tuning of bacterial motility by mutations located in the message segment of Pa flg22, Related to Figure 2.5. A and B- Pictures showing the motility phenotype of *Pa* Δ*fliC* **(A)** and *Ps* Δ*fliC* **(B)** strains complemented by either an empty vector (EV) or with *fliC* variants coding the flg22, flg22^{D14K}, flg22^{I21D}, flg22^{Q20D}, flg22^{G18F} and flg22^{I21E} epitopes. Scale bar: 0.5 cm. **C-** Pictures showing the motility phenotype of *Ps* Δ*fliC* strains complemented by either an empty vector (EV) or with *fliC* variants coding the flg22, flg22^{A16D}, flg22^{K13D}, flg22^{S11D}, flg22^{A22D} flg22^{K13Y} and flg22^{A22E} epitopes. Scale bar: 0.5 cm. **D-** Quantification of swimming motility. The box plot represents the first and third quartiles, centered by the median. Whiskers include the 10th-90th percentile of the data points. The boxplots are marked with the following color code: Pa flg22 (in orange), all others (purple). Statistical significance was assessed using one-way ANOVA followed by Tukey's HSD for all pairwise comparisons ($P < 0.05$). Genotypes with the same letter are indistinguishable at >95% confidence. The number of biologically independent observations (n) is indicated on top of the graph. The photos in (A) were taken with the plates closed and the halo reflection on the lid is visible for flg22^{D14K}, flg22^{Q20D} and flg22^{I21E}.

A

flg22 variant tested	k_a [$M^{-1}s^{-1}$]	k_d [s^{-1}]	K_D [μM]
Pa flg22	7.57 x 10^3	3.6 x 10^{-3}	0.48
flg22^{A16D}	5.89 x 10^3	5.59 x 10^{-3}	0.95
flg22^{K13D}	5.07 x 10^3	2.79 x 10^{-3}	0.55
flg22^{S11D}	7.31 x 10^3	6.48 x 10^{-3}	0.89
flg22^{A22D}	6.86 x 10^3	5.23 x 10^{-3}	0.76
flg22^{K13Y}	5.95 x 10^3	4.87 x 10^{-3}	0.82
flg22^{A22E}	6.23 x 10^3	5.52 x 10^{-3}	0.88
flg22^{I21D}	1.15 x 10^4	7.1 x 10^{-2}	6.2
flg22^{Q20D}	3.34 x 10^3	1.87 x 10^{-2}	5.6
flg22^{G18F}	5.44 x 10^4	8.49 x 10^{-2}	1.6
flg22^{I21E}	5.63 x 10^3	3.24 x 10^{-2}	5.8

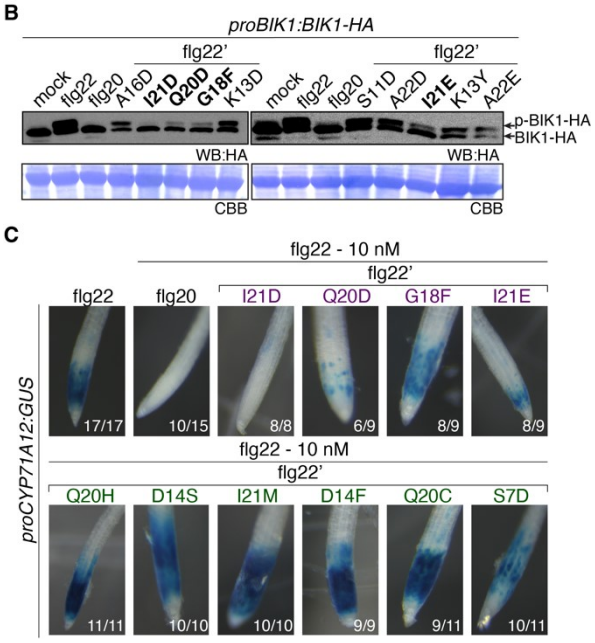


Figure 2.13. niSEI peptides fail to induce BIK1 phosphorylation and antagonize the induction of *CYP71A12* by Pa flg22, Related to Figure 2.6. A- Table summarizing the Grating-coupled interferometry (GCI)-derived binding kinetics for flg22-bound FLS2 and BAK1. The corresponding flg22 variants, association (k_a) and dissociation (k_d) rate constants and the dissociation constant K_D are indicated. The sensorgrams with raw data and the respective fits are presented in (Data S4B). **B-** Western blot analyses of BIK1 phosphorylation in Arabidopsis transgenic seedlings (*proBIK1:BIK1HA)* either untreated (mock) or treated for 10 min with different flg22 peptides. The identity of the peptides is shown on top. niSEI peptides are marked in bold. anti-HA antibodies were used to analyse the lysates. This experiment was repeated three times with similar results. **C-** Root expression patterns of the defense marker gene *CYP71A12* using a beta-glucuronidase transcriptional reporter system. Responses to Pa flg22 alone or co-treated with 100 µM of the epitopes were analyzed. Shown are representative images of the root. The epitopes are indicated on top of each panel. The number of roots that showed a phenotype corresponding to the representative image compared to the total number of roots tested is indicated in the picture.

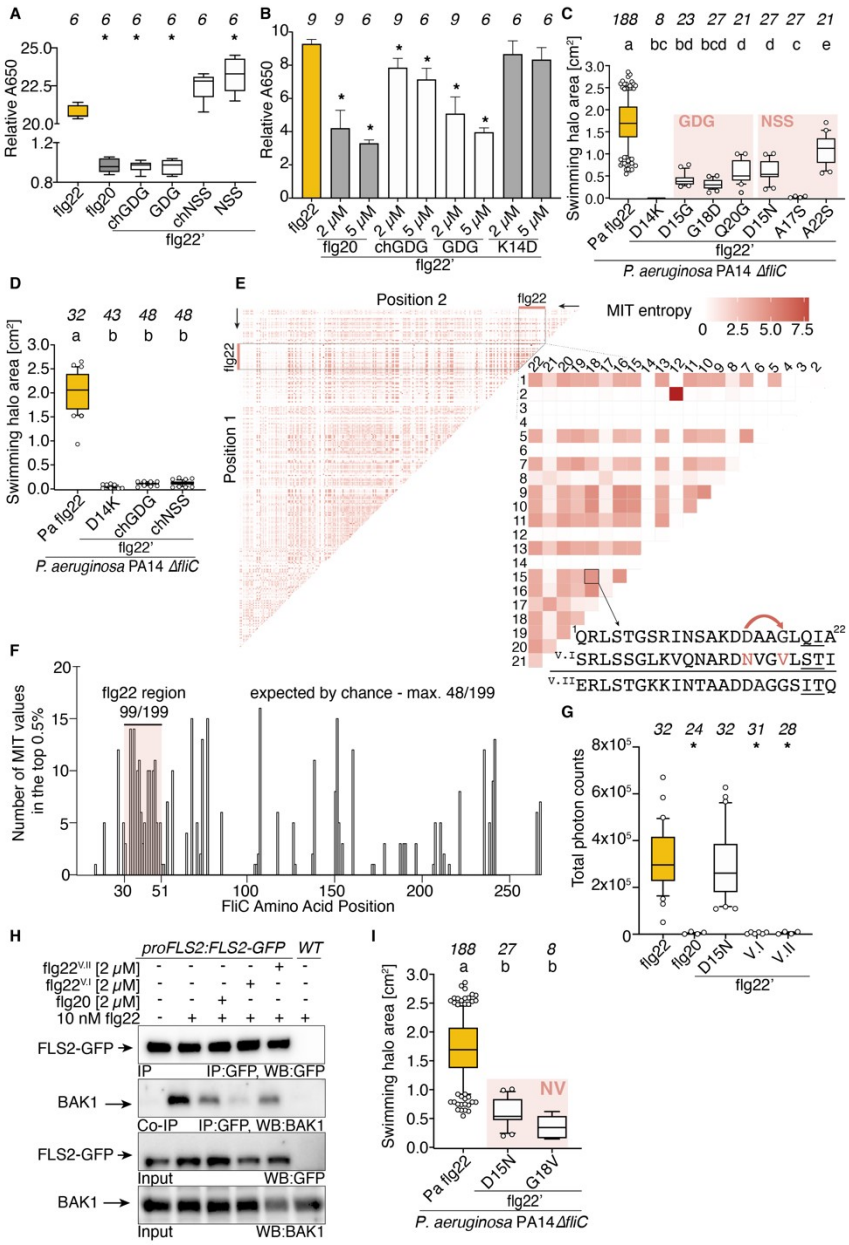


Figure 2.14. Arabidopsis-specific lineages of *Pseudomonas* encode receptor antagonists, Related to Figure 2.7. A and B- FLS2-BAK1 heterocomplex formation measured *in vitro* using an enzymelinked immunosorbent assay. **A-** Plotted is the relative absorbance (A650 nm) after 2 h obtained in the presence of 100 nM of the peptide indicated at the bottom. **B-** Represented is the relative absorbance (A650 nm) after 2 h obtained in the presence of 10 nM Pa flg22 peptide and different concentrations of flg20, flg22^{chGDG}, flg22^{GDG} or flg22^{D14K}. The error bars indicate the standard deviation from at least two biological repetitions. **C and D -** Quantification of swimming motility for *Pa* D*fliC* strains complemented with *fliC* variants coding either flg22, flg22^{D14K} or the indicated variants. **E-** Mutual information theory (MIT) heat map of FliC proteins sequences extracted from *Ps* OTUs (Karasov et al., 2018). The flg22 segment is highlighted and magnified on the right. The colour scale bar on top shows the MIT entropy score values for position pairs. The arrows on the flg22 sequence at the bottom indicate that Asn^{15} co-

varies with Val^{18} in the Var I epitope (V.I). The amino acid sequence below defines the Var II epitope (V.II). **F-** Distribution of the positions in the top 0.5% MIT values across the FliC of *Ps* OTUs. The flg22 segment is highlighted in red. **G-** Induction of oxidative bursts represented as total photon counts over 40 mins for Pa flg22, flg20, $flg22^{D15N}$, Var I ($flg22^{V.I}$) and Var II ($flg22^{V.II}$). **H-** Western blot analyses of FLS2–BAK1 co-immunoprecipitations (Co-IP/IP) in Arabidopsis transgenic seedlings (*proFLS2:FLS2-GFP)* either untreated (mock) or treated for 10 min with the flg22 peptides indicated on the left at the concentrations indicated on top. anti-BAK1 or antiGFP antibodies were used to analyze the lysates. This experiment was repeated two times with similar results. **I-** Quantification of the swimming motility of *Pa* D*fliC* strains complemented with *fliC* variants coding $flg22^{D15N}$ or $flg22^{G18V}$. Both experiments were repeated two times with similar results. **A, C, D, G and H-** The box plot represents the first and third quartiles, centered by the median. Whiskers include the 10^{th}-90^{th} percentile of the data points from two (A, H) and three (C, D, G) independent biological experiments. **A, B and H-** Statistical significance was assessed using one-way ANOVA followed by Dunnett's multiple comparison test (*$P < 0.05$). **C, D, I-** Statistical significance was assessed using one-way ANOVA followed by Tukey's HSD for all pairwise comparisons ($P < 0.05$). Genotypes with the same letter are indistinguishable at >95% confidence. The number of biologically independent replicate tested (n) is indicated on top of each graph.

Section 2.6: Methods

Materials Availability. Plasmids and bacterial strains generated in this study will be made available on request, but we may require a completed Materials Transfer Agreement if there is potential for commercial application.

Data and Code Availability. Original/source data for flg22 peptide array and Pseudomonas mutation scans for bacterial swimming motility is available at: https://github.com/kateparys/Parys-et-al.-2021. The codes and raw data for Mutual Information Theory is available at https://github.com/ncolaian/flagellin_antag_pleiotropy.

Plant material and growth conditions. *Arabidopsis thaliana* accession Columbia (Col-0) was used as wild type in all experiments performed. The mutant line *fls2 efr* (SAIL_691_C4 / Salk_044334) was described elsewhere (Nekrasov et al., 2009). The transgenic lines *proFLS2:FLS2-GFP* (Chinchilla et al., 2007), *proMYB51:NLS-3mVenus, proWRKY11:NLS-3mVenus*, *proCYP71A12:GUS* (Poncini et al., 2017), and *proBIK1:BIK1-HA* (Lin et al., 2014) were described previously. Plants were grown on soil in short- day light conditions (12 h light/12

h dark) or vertically in Petri dishes containing 1/2 Murashige and Skoog (MS) medium, 0.8% plant agar, and 1% sucrose in long-day light conditions (16 h light/8 h dark).

Deletion of *fliC* from *Pseudomonas aeruginosa (Pa)* PA14. To construct an in-frame deletion of *fliC,* two PCR products flanking the *fliC* gene were obtained from *Pa* PA14 chromosomal DNA by PCR amplification with primer pairs O183/P183 and Q183/R183 (Table S6) (Liberati et al., 2006). The combined 729-bp upstream and 723-bp downstream fragments were used as a template for a second overlapping PCR with primers O183 and R183, because primers P183 and Q183 contained a complementary sequence. The resulting fragment, a 1464-bp deletion spanning the entire *fliC* coding region except for the stop codon (coordinates 4465410-4466873 of the *Pa* PA14 genome (Winsor et al., 2016)), was digested with BamHI and KpnI and ligated into the corresponding sites of the suicide vector pEXG2 (Rietsch et al., 2005), resulting in plasmid pEXG2-Δ*fliC*. All DNA manipulations were verified by DNA sequencing. The strain *Pa* PA14 *ΔfliC* was constructed by homologue recombination. Briefly, plasmid pEXG2-Δ*fliC* was mobilized into the strains *Pa* PA14 with the aid of *E. coli* strain S17-1, and then chromosomally integrated through selection for gentamicin. Excision of the vector by a second crossover was achieved by selection of sucrose insensitive cells as the pEXG2 vector encodes the *Bacillus subtilis* sacB gene, whose gene product - levansucrase - renders Gram-negative cells sensitive to sucrose (Hmelo et al., 2015; Rietsch et al., 2005).

Cloning of allelic variants of *Pa fliC* gene and complementation *of Pa ΔfliC.* *Pa fliC* gene (NZ_JTUE01000011.1) was designed to comprise the FliC coding sequence and regions 405 bp upstream of the start codon and 78 bp downstream of the stop codon. Using gene synthesis (Genewiz®), the *fliC* gene was cloned into the modified broad-host-range vector pBBR1-MSC5 (Kovach et al., 1995) using the BstBI and NsiI restriction sites. The mutant library of 412 *fliC*

allelic variants was then constructed using the same method. The resulting pBBR1-MSC5-*fliC*' plasmids were verified by DNA sequencing and introduced into *Pa ΔfliC* via conjugation. Briefly, overnight liquid cultures of the *E. coli* helper strain pRK2013, *Pa* Δ*fliC*, and *E.coli* transformed with pBBR1-MSC5-*fliC*' plasmids were mixed in a ratio 1:1:1 and spotted on a Luria broth (LB) plates (Sambrook et al., 1989). The next day, positive transformants were selected on LB supplemented with gentamycin [25 □g/ml] and chloramphenicol [10 □g/ml]. Positive transformants were verified by colony PCR using the fliC_pBBR1-MSC5_FP and fliC_pBBR1- MSC5_RP primers (Table S6). 20 randomly selected transformants were subjected to Sanger sequencing for verification of inserts.

Quantification of bacterial motility in swim plate assay. *Pa* Δ*fliC* strains were grown overnight on LB agar plates supplemented with appropriate antibiotics at 37□C. The next day, fresh swim plates were prepared with M9 Minimal media supplemented with 0.3% agar (Ha et al., 2014; Sambrook et al., 1989). The plates were then tooth-pick inoculated with a sterile 2 µl pipette tip by placing the tip in an overnight culture and then onto the swim plate. Pictures of the plates were acquired 24 hours after inoculation. The diameter (d) of the bacterial halo was measured in Image J and the halo area was calculated using the formula $(\pi/4) \times d^2$. To confirm the expression of flagellin in *P. aeruginosa fliC* allelic clones with reduced or no motility, we performed Western Blot analysis using Anti-flagellin antibodies (Figure 2.8B). We suspect that the molecular shift of the band corresponding to flagellin (Fig. S1) may be caused by different flagellin glycosylation patterns and/or other reasons. The swimming motility assay for *Pseudomonas syringae* pv *tomato* DC3000 *(Ps)* Δ*fliC* mutants was performed as previously described (Clarke et al., 2013). Briefly, the bacteria were toothpick inoculated on freshly prepared 0.25% agar plates containing King's B medium (KB) supplemented with appropriate

antibiotics (King et al., 1954). The plates were incubated at 28□C and pictures of bacteria were acquired 48 hours after inoculation. The diameter (d) of the bacterial halo was measured in Image J and the halo area was calculated using the formula $(\pi/4) \times d^2$.

FLS2 extracellular domain expression and purification for peptide array interrogation. The extracellular domain (ECD) of FLS2 from *Arabidopsis thaliana* (amino acids 24 - 804) was cloned into the baculovirus transfer vector pMeI Bac B1 (Invitrogen). A C-terminal Strep II - 6xHis tag was fused to the ECD. The $FLS2^{LRR}$ - StrepII-6xHis ECD was produced by secreted expression in baculovirus-infected High Five insect cells and harvested 72 hours post-infection. Subsequently, the protein was purified using Ni-NTA chromatography (Ni Sepharose excel, GE Healthcare) and subjected to size-exclusion chromatography (Superdex 200 16/60, GE Healthcare) in buffer containing 10 mM Bis-Tris pH 6.0, 100 mM NaCl. The protein was concentrated to ~ 3.0 mg/ml.

High-density peptide array design and analysis of interaction scores. The peptide microarrays were made by and purchased from PepperPrint® (Beyer et al., 2007; Stadler et al., 2008). The custom arrays were designed to comprise single amino acid mutational scans of both the full-length Pa flg22 peptide (QRLSTGSRINSAKDDAAGLQIA) and epitope variants truncated at either or both of the N- and C- terminal moieties through a stepwise deletion series from both extremities of Pa flg22 (Figure 2.9A). All flg22 peptides were printed at least in duplicate, while the non-mutated reference variants were printed 23 times on the array. Overall, 3,268 flg22 peptides were interrogated with the purified $FLS2^{ECD}$ at pH 6 and with a protein concentration of 20 µg/ml, followed by staining with the secondary anti-6xHis Epitope Tag DyLight 680 antibody. HA and Flag control peptides framed the array and served as an internal quality control. The on-array FLS2-flg22 interactions were quantified using LI-COR Odyssey

Imaging System, based on the 16-bit gray scale tiff files at scanning intensities of 7/7 and the microarray image analysis was performed using PepSlide® Analyzer. Given that different signal intensities were observed for the different peptide lengths, the data was categorized into four subgroups (Figure 2.9A). To calculate the interaction score, the raw median intensities for each peptide were Log2 transformed to ensure symmetric distribution of the data. To assess the internal consistency of the flg22 array data, correlation between the three sectors of the full-length substitution scan of flg22 was calculated using Spearman's correlation coefficient ($r \sim 0.7 - 0.9$). Next, we applied a Median Absolute Deviation (MAD) scale normalization to calculate the interaction scores, where the unmutated variant of the flg22 derivative for each category was established as a reference sample for a respective dataset (Kappal, 2019; Leys et al., 2013). The residuals from the non-mutated Pa flg22 (or truncated variants, respectively) printed at random at least 23 times were used to calculate the median MAD for each subgroup. Subsequently, modified Z-scores for $FLS2^{ECD}$ interactions were calculated for each array spot using their respective non-mutated flg22 variant's median and median absolute deviation with a formula: Z-score = (Log2 median $intensity^{flg22'\ peptide}$ - Log2 median $intensity^{Pa\ flg22}$)/$MAD^{Pa\ flg22}$.We then calculated the median Z-score for each individual peptide variant, hereafter referred to as the interaction score. To understand the variability between interaction scores for peptides derived from the full-length Pa flg22 substitution scan, we applied descriptive statistics. Flg22 peptides with an interaction score above or below the canonical Pa flg22 interquartile region (IQR) were assigned as variants with strong/weak on-chip interaction with $FLS2^{ECD}$. Variants that reside outside of the 1.5*IQR of Pa flg22 (interaction score > 3.12 and < -2.86, respectively) were considered to be significant on-chip $FLS2^{ECD}$ binders/non-binders for the purpose of functional validation of the peptide array data.

Design and cloning of GreenGate compatible, fluorophore tagged broad host range vector. To quantify the motility of *Ps* Δ*fliC* strains, we generated the vector pGGrhizo_Plum. pGGrhizo_Plum was designed to be compatible with modules for the green-gate modular cloning toolkit (Lampropoulos et al., 2013) and thus uses the same sticky-end overhangs as green-gate modules A-F. In contrast to green-gate, there is no resistance cassette module (module G), because the vector backbone carries a gentamycin resistance gene (*gmR*). The green-gate acceptor module was cloned from pGGZ002 using primers 13 and 14 (Supplemental Table 6). The Plum expresses mPlum fluorescent protein (primers 13-18), driven by *lacP*. The mPlum, replication origins and resistance cassettes were amplified from rainbow vectors (Barbier and Damron, 2016) and assembled with the greengate cloning site using Gibson assembly (Gibson et al., 2009). Inserts were verified by sanger sequencing using primers 21 and 22 (Supplemental Table 6).

Cloning and complementation of *Ps fliC*. The *Ps fliC* gene, including its promoter and terminator regions (Clarke et al., 2013), was PCR amplified from genomic DNA using Phusion High-Fidelity DNA Polymerase (Thermo Scientific, Cat. No. F530S) and primers fliC_Ps.syr_GreenGate FP/RP (Supplemental Table 6). To generate the empty vector (EV) control, primers fliC_Ps.syr_GreenGate FP and fliC_Ps.syr_GreenGate_EV were used. The PCR products were separated via gel electrophoresis and purified using the Wizard® SV Gel and PCR Clean-Up System (Promega). The insert was cloned into the broad-host-range pGGrhizo_Plum vector using Gibson assembly. Correct clones were verified by colony PCR and Sanger sequencing using primers listed in Supplemental Table 6. Mutations (single or multiple) within the flg22 segment of the *fliC* gene were introduced via PCR-based site-directed mutagenesis using Phusion Flash High-Fidelity polymerase (Thermo Scientific, Cat. No. F548S) and primers

listed in the Supplemental Table 6. The PCR products were separated via gel electrophoresis, purified from agarose gel, subjected to restriction digest with FastDigest DpnI restriction enzyme (Thermo Scientific, Cat. No. FD1704) and transformed into NEB 10-beta competent *E. coli* (NEB, Cat. No. C3019I). The resulting pGGrhizo_Plum -*fliC*" plasmids were verified by DNA sequencing and introduced into *Ps* Δ*fliC* via conjugation as described above. Positive transformants were selected on KB agar plates supplemented with gentamycin [25 □g/ml], chloramphenicol [25 □g/ml] and rifampicin [25 □g/ml] and verified by colony PCR using fliC_syr_FP_339 and Ps.syr-qPCR_RP primers (Supplemental Table 6).

Microscopic observation of bacterial motility in liquid media. Liquid cultures of *Ps* Δ*fliC* strains grown overnight in KB media at 28□C were diluted to 600 nm OD_{600} = 0.15. 15 µl of bacterial suspension was loaded onto µ-Slide VI0.4 (Ibidi, Cat. No. 80606) channels and examined using a Zeiss AXIO Vert A1 with phase contrast filter. Movies recording bacterial movement in liquid media were obtained from 200 images using SPOT 5.1 software.

Expression and purification of $FLS2^{ECD}$ and $BAK1^{ECD}$ for GCI assay. The ECDs of FLS2 (amino acids 1 – 800) and BAK1 (amino acids 1 – 220) were amplified from *A. thaliana* genomic DNA or cDNA and cloned into a modified pFastBac vector (Geneva Biotech), containing a TEV (tobacco etch virus protease) cleavable C-terminal StrepII - 9xHis tag as well as a non-cleavable Avi tag (Cull and Schatz, 2000; Fairhead and Howarth, 2015) for FLS2. *Trichoplusia ni* (strain Tnao38 (Hashimoto et al., 2010)) cells were infected with a multiplicity of infection (MOI) of 1 at a density of $2x10^6$ cells/ml and incubated for 26 h at 28°C and 48 h at 22°C. The secreted protein was purified using Ni-NTA (HisTrap excel; GE Healthcare; equilibrated in 25 mM KPi pH 7.8, 500 mM NaCl) and StrepII (Strep-Tactin Superflow high capacity; IBA; equilibrated in 25 mM Tris pH 8.0, 250 mM NaCl, 1 mM EDTA) affinity

chromatography. The StrepII – 9xHis tag was cleaved by incubation with a His tagged TEV protease at 4°C and removed through a second Ni^{2+} affinity chromatography step. Protein purity was further improved by size-exclusion chromatography on a HiLoad 16/600 Superdex 200 column (GE Healthcare), equilibrated in 20 mM sodium citrate pH 5.0, 250 mM NaCl. His-tagged BirA was purified from *E. coli* by Ni^{2+} affinity chromatography.

Biotinylation of $FLS2^{ECD}$. Purified $FLS2^{ECD}$ containing the Avi tag (20 – 100 μM) was incubated with BirA (Fairhead and Howarth, 2015) for 1 h at 25°C in a total volume of 200 μl 25 mM Tris pH 8, 150 mM NaCl, 5 mM $MgCl_2$, 2 mM 2-Mercaptoethanol, 0.15 mM Biotin, 2 mM ATP. BirA as well as free biotin were subsequently separated from the biotinylated FLS2 ectodomain through size-exclusion chromatography.

Peptide synthesis. Pa flg22, flg20 and flg22 peptide derivatives (Data S1) were synthesized by our in- house protein chemistry facility at >95% purity or purchased from Shanghai Apeptide Co. An additional leucine (L) was added to the N-terminus of all peptides to avoid conversion of glutamine (Q) to pyroglutamate. The peptides were dissolved in pure water to a final concentration 100 □M, aliquoted, and stored at -20°C. Peptide purity was validated by MALDI-TOF MS (Data S2).

Grating coupled interferometry (GCI). GCI experiments were performed on a Creoptix WAVE system (Creoptix AG, Switzerland), a label-free surface biosensor (Kozma et al., 2009), using 4PCP WAVEchips (quasi-planar polycarboxylate surface; Creoptix AG, Switzerland). After a borate buffer conditioning (100 mM sodium borate pH 9.0, 1 M NaCl; Xantec, Germany) strepdavidin was immobilized through a standard amine coupling protocol, followed by passivation of the surface (0.5% BSA [Roche, Switzerland] in 10 mM sodium acetate pH 5.0) and final quenching with 1 M ethanolamine pH 8.0 for 7 min (Xantec, Germany). The

biotinylated $FLS2^{ECD}$ was coupled to the strepdavidin coated chip until the desired density was reached. For peptide binding experiments, the respective peptide was injected in a 1:2 dilution series in 20 mM citrate pH 5.0, 250 mM NaCl at 25°C. BAK1 binding kinetics were determined in a 1:2 dilution series, with the respective peptide present at 0.5 □ *K*D in all blank - and BAK1 injections. Blank injections were used for double referencing and a DMSO calibration curve was used for bulk correction. Analysis and correction were performed using the Creoptix WAVEcontrol software (applied corrections: X and Y offset, DMSO calibration, double referencing) using a one-to-one binding model.

Plant material and growth conditions. *Arabidopsis thaliana* accession Columbia (Col-0) was used as wild type in all experiments performed. The mutant line *fls2 efr* (SAIL_691_C4 / Salk_044334) was described elsewhere (Nekrasov et al., 2009). The transgenic lines *proFLS2::FLS2-GFP* (Chinchilla et al., 2007), *proMYB51::NLS-3mVenus, proWRKY11::NLS-3mVenus*, *proCYP71A12::GUS* (Poncini et al., 2017), and *proBIK1::BIK1-HA* (Lin et al., 2014) were described previously. Plants were grown on soil in short-day light conditions (12 h light/12 h dark) or vertically in Petri dishes containing 1/2 Murashige and Skoog (MS) medium, 0.8% plant agar, and 1% sucrose in long-day light conditions (16 h light/8 h dark).

MAMP responses assays. ROS bursts and seedling growth inhibition (SGI) assays were performed as previously described (Lozano-Duran and Belkhadir, 2017). Briefly, for ROS burst assays, leaf discs (diameter 5 mm) were cut out from 6-week-old healthy Col-0 or *fls2 efr* plants (at least 12 discs per peptide tested). The discs were placed in a 96-well plate (Griner Bio one, Cat. No. 675 074) containing 100 µl sterile MonoQ water, with the adaxial side up. Discs were vacuum infiltrated for 10 min and incubated for 5 h in continuous light at 21□C. Alternatively, when using Sigma-Aldrich luminol at a concentration 0.034 mg/ml (see below), the discs were

left in the dark for 16 hours, (Supplemental Figure 2.4F). Next, the water was carefully removed and replaced with 100 µl of elicitation solution containing HRP (Sigma- Aldrich) at a final concentration of 0.04 mg/ml or 0.02 mg/ml (Fig. S9B), L-012 at a final concentration of 0.068 mg/ml or 0.034 mg/ml (Wako Chemicals GmbH or Sigma-Aldrich, respectively), and flg22 peptide at concentrations ranging from 10 nM to 10 □M. Immediately afterwards, the measurement of relative luminescence was started using a BiTec Synergy 4 microplate reader. The luminescence was measured every minute for 40 min and the total relative light units for the 40 time points were used for the analysis. For the measurement of SGI, Col-0 or *fls2 efr* mutant plants were grown vertically on 1/2 MS plates for 5 days. Next, the seedlings were transferred to a 48-well plate (Griner Bio one, Cat. No. 677180) containing 1 ml of liquid MS medium supplemented with flg22 peptides at a concentration of 10 nM. The fresh weight of each seedling was measured 9 days later. Each experiment included the mock control and seedlings treated with flg20 as a negative control. At least 8 seedlings per treatment were measured. The inhibition of seedling growth was calculated for each flg22 peptide compared to the median fresh weight of the mock control.

Plant bacterial assays. *Pseudomonas syringae pv. tomato* DC3000 ΔfliC - EV, -FliC Pa flg22 and -FliC Pa flg22^{Q20D} were performed as previously described (Zipfel et al., 2014) with some modifications. Bacterial growth in plant leaves was assessed by inoculating 4-week-old WT and *fls2 efr* plants with a bacterial inoculum of OD600 = 0.2 in 10 mM MgCl2 supplemented with 0.02 % Silwet L-77. Three days after inoculation for 8 leaf discs from 4 different plants for each sample were bored from the infiltrated area and homogenized in 10 mM MgCl2. The bacterial titer was determined by plating and serial dilution. Disease symptoms were scored, and pictures of the leaves were taken, five days after inoculation.

Imaging of fluorescent transcriptional reporter lines. 5-day-old seedlings grown on 1/2 MS plates were transferred to liquid medium supplemented with the different flg22 peptides at 10 nM. After 24 hr, confocal laser scanning microscopy (CLSM) experiments were performed using a Zeiss LSM780 microscope. Excitation and detection windows were set at 488 nm and 493-598 nm respectively. The adaxial side of the leaves for *proWRKY11::NLS-3mVenus* and root tip for *proMYB51::NLS-3mVenus* were imaged. Confocal images were processed and analyzed using (Schindelin et al., 2012).

Imaging of the CYP71A12 beta-glucuronidase transcriptional reporter system. 5-8 *proCYP71A12::GUS* seedlings were germinated in a 48-well plate (Griner Bio one, Cat. No. 677180) in 1/2 MS liquid supplemented with 0.5 g/l MES (Millet et al., 2010). After 7 days, a solution of antagonist peptide at a final concentration 100 µM was added and pre-incubated with seedlings for 15 minutes. Next, the media was supplemented with 10 nM Pa flg22 and incubated 5 hours at 21□C. The media was then removed, and each well was washed with 50 mM sodium phosphate (pH 7). The GUS substrate solution was added (50 mM Na3PO4, 10 mM EDTA, 0.5 mM K4[Fe(CN)6], 0.5 mM K3[Fe(CN)6], 0.5 mM X-Gluc, 0.01 % Silwet L-77) and incubated overnight in the dark at 37□C. Seedlings were then fixed in 3:1 EtOH:acetic acid at 4□C for 5 hours and stored in 95% EtOH. Root pictures were taken using a Leica M205FA stereoscope coupled to a Leica DFC310FX camera.

FLS2-BAK1 interaction using an using an enzyme-linked immunosorbent assay. Ectodomain interaction studies were performed as previously described (Smakowska-Luzan et al., 2018) with minor modifications. $FLS2^{ECD}$ cloned in pECIA2 (bait protein) and $BAK1^{ECD}$ cloned in pECIA14 vector (prey protein) were expressed via transient transfection in *Drosophila melanogaster* Schneider 2 (S2) cells with Expres2 TR Transfection Reagent (Expression

System). Upon transfection, the cells were shifted from 27°C to 21°C. 24 h later, protein expression was induced with 1 mM $CuSO_4$. Supernatant was collected three days after induction and protein expression was confirmed by immunoblotting using anti-V5 (Invitrogen, Cat. No. R961-25) and anti-Flag (Sigma Aldrich, Cat. No. A8592) antibodies to detect the bait and prey proteins respectively. $FLS2^{ECD}$ and $BAK1^{ECD}$ were diluted in 1x PBS containing 0.1 % Tween-20 (PBS-T) and mixed 1:1. flg22 peptide was added to mix to a final concentration ranging from 10 nM – 1 µM, depending on the assay. The protein mix was pre-incubated 2 h at 4□C. 100 µl of the protein solution was then transferred to a protein-A coated 96 well plate (Sigma Aldrich, Cat. No. 15132) and incubated overnight at 4°C. The next day, the plate was washed two times with 100 µl PBS-T and 100 µl of alkaline phosphatase was added (Sera care, Cat. No. 5120-0059). Two hours after addition of the substrate, absorbance was measured at 650 nm using a Synergy H4 Multi-Mode plate reader (BioTek). Each plate contained a BIR4-BAK1 positive control (Smakowska-Luzan et al., 2018). To control for unspecific binding of prey protein to protein A coated wells, the bait protein was substituted with a solution of human recombinant IgG1-Fc protein (Invitrogen, Cat. No. 13445946) at a final concentration of 625 pg/µl. For each flg22 peptide tested, relative A650 was quantified by dividing the raw A650 after 2 h to the signal of the mock control.

Protein extraction and co-immunoprecipitation in Arabidopsis. 0.8 g of 2-week-old seedlings of *proFLS2::FLS2-GFP* or Col-0 grown vertically on 1/2 MS plates were placed in a 6-well plate (Griner Bio one, Cat. No. 657185) and rinsed with 5 ml of sterile water. Next, the water was exchanged for either elicitor solution at final concentration of 100 nM, or with 10 nM Pa flg22 and increasing concentrations of antagonistic flg22 peptide. The seedlings were vacuum infiltered for 10 min and immediately afterwards ground in liquid nitrogen to obtain a fine powder.

2-3 ml of extraction buffer (50 mM HEPES pH 7.5, 100 mM NaCl, 10 mM MgCl2, 10% glycerol, 2 mM EDTA, 1 mM Na2MoO4, 20 mM NaF, 1 mM DTT, EDTA-free protease inhibitor cocktail (Roche, Cat. No. 45-5056489001), 1% IGEPAL) was added to each sample. The tissue was additionally homogenized with a Polytron (Kinematica) and incubated for 30 min at 4□C. The samples were then centrifuged for 10 min, 4□C, 13000 rpm and the supernatant were filtered through Miracloth (Merck Millipore Calbiochem, Cat. No. 475855-1). Next, the supernatant was diluted 1:1 in extraction buffer without IGEPAL, and 15 µl of equilibrated GFP-Trap A beads (Chromotek, Cat. No. gta-100) were added to each sample. The samples were incubated for 4 h at 4□C with rotation (12 rpm) and washed 4 times with 1 ml of washing buffer (50 mM HEPES pH 7.5, 100 mM NaCl,10 mM MgCl2, 0.1% IGEPAL). The protein complexes were eluted with Laemmli sample buffer (2x), boiled for 10 min at 95□C and analyzed by SDS-PAGE followed by immunoblotting using anti-GFP-HRP (Sigma) and anti-BAK1 (Agrisera, Cat. No. AS12 1858) antibodies.

BIK1 phosphorylation assay. Four 2-week-old *proBIK1::BIK1-HA* seedlings (Kadota et al., 2014), grown on ½ MS plates, were transferred to a 6-well plate (Griner Bio one, Cat. No. 657185) and rinsed with sterile water. Next, the water was replaced by a solution containing flg22 at a final concentration of 100 nM. The seedlings were vacuum infiltrated for 10 min. Immediately afterwards, the seedlings were transferred to 2 ml tubes containing glass beads and 300 µL of protein extraction buffer (50 mM TRIS pH 7.5, 100 mM NaCl, 10 % glycerol, 5 mM EDTA, 1 mM Na2MoO4, 20 mM NaF, 1 mM DTT, EDTA-free protease inhibitor cocktail (Roche, Cat. No. 45-5056489001)). The samples were ground with a Precellys homogenizer (2 x 40 sec; 4800 rpm). After centrifugation (10 min, 13000 rpm, 4ºC), the supernatant was boiled for 10 min at 95°C with Laemmli sample buffer. Analysis of BIK1 phosphorylation was performed

by SDS-PAGE followed by immunoblotting using anti-HA (Sigma Aldrich, Cat. No. A8592) antibodies.

Analysis of co-mutation sites in FliC. Mutual Information Theory (MIT) was used to identify residues in FliC proteins that are coevolving in *Pseudomonas* (Gouveia-Oliveira and Pedersen, 2007). Two datasets, 528 complete genomes from the online Pseudomonas Database (Winsor et al., 2016) and 1490 genomes (Karasov et al., 2018) were downloaded and FliC proteins were identified as described in (Colaianni et al., submitted). We then used muscle to align the FliC proteins, and trimal to remove columns with gaps in 10% or more of the sequences for each dataset separately (Capella-Gutiérrez et al., 2009; Edgar, 2004). We calculated the Mutual Information (MI) and performed row-column weighting (RCW) as described in (Gouveia-Oliveira and Pedersen, 2007). Briefly, MI was calculated with the equation:

$$MI(A;B) = \sum_i \sum_j P(a_i,\ b_j) log \left[\frac{P(a_i,\ b_j)}{P(a_i)P(b_j)} \right]$$

A and *B* are the sites being compared, and *i* and *j* run through all of the amino acids at the site. The estimation of the probabilities was performed by using the observed frequencies found in the dataset. After calculating MI at each site, we performed RCW to adjust the MI for sites that are rapidly evolving. This increases the chances of identifying true co-evolving sites. We used the formula in (Gouveia-Oliveira and Pedersen, 2007)

$$RCW(A;B) = \frac{MI_{ij}}{\frac{MI_{.i} + MI_{.j} + MI_{j.} + MI_{i.} - 2MI_{ij}}{2n-2}}$$

Where $MI_{.j}$ denotes the sum of the Mutual Information matrix over all lines in column *j.* These adjusted MI values were then used to calculate Z-scores for each value. We considered the residue pairs with MI values in the top 0.5% as co-evolving sites.

Identification of non-synonymous SNPs in natural accessions of Arabidopsis. To

parse the non- synonymous single nucleotide polymorphisms (SNPs) in *FLS2* and *BAK1* from the genomes of 1135 naturally inbred lines of *Arabidopsis thaliana* (Alonso-Blanco et al., 2016b), we build an in-house snprator.r script in the R programming environment. The script and necessary genomic data are freely accessible under the link: https://github.com/arthurkorte/SNP_extractor. As an output, a list of identified SNPs, the chromosomal positions, frequency of the SNPs, and accession ID numbers that carry the SNP are retrieved (Table S6).

Programs used for statistical analysis and data visualization. GraphPad PRISM 8.0 and R programming environment (version R 3.6.0) were used for data analysis and visualization. In all assays performed, data was combined from at least two independent biological experiments. Unless specified otherwise, the box plot represents the first and third quartiles, centered by the median. Whiskers include the 10^{th}-90^{th} percentile of the data points. Dots represent individual observations. Outliers were removed using the ROUT test, implemented in GraphPad PRISM (Q = 0.1%). Non-mutated Pa flg22 peptide was used as a control sample and comparator in all statistical tests performed. Unless otherwise indicated, the calculated P-values indicate the likelihood that the tested variant differs significantly from Pa flg22 at a 95% confidence level. One-way ANOVA followed by Tukey's multiple comparisons test was performed using GraphPad PRISM 8.0. The ROS burst and SGI results were analyzed using a mixed effect linear model as previously described (Smakowska-Luzan et al., 2018). Briefly, the statistical analyses were performed in R version 3.6.0. The lme4 package was used for model construction. In this model, the non- mutated flg22 was used as a fixed effect and each experiment performed as a random effect using the formula ROS ~ Variant + (1 | Trial) and SGI ~ Variant + (1 | Trial), respectively. Subsequently, the lmerTest package was implemented to

calculate p-values using Satterthwaite approximation and the resulting P values were corrected for multiple testing using the Holm method. To determine if peptides produced significant ROS and SGI phenotypes (Figure 2.11C), we compared the 95% confidence intervals for the mean of each peptide's measurements to the 95% confidence intervals for the mean in 10000 independent samples with equal n from the Pa flg20 measurements**.** A significant difference was designated when the confidence intervals did not overlap in >95% of the samples. To estimate the relationship between bacterial motility and FLS2 binding within different categories designated based on the binding affinities, Kolmogorov-Smirnov test in R programming environment was used (R version 3.6.0). Flagellin structure was modeled using PyMOL software (PyMOL v1.6.0.0) and flg22 amino acid sequence LOGOs were generated using WebLOGO portal (Crooks et al., 2004).

REFERENCES

Alonso-Blanco, C., Andrade, J., Becker, C., Bemm, F., Bergelson, J., Borgwardt, K.M., Cao, J., Chae, E., Dezwaan, T.M., Ding, W.*, et al.* (2016b). 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. Cell *166*, 481-491.

Barbier, M., and Damron, F.H. (2016). Rainbow Vectors for Broad-Range Bacterial Fluorescence Labeling. PLoS One *11*, e0146827

Bardoel, B.W., van der Ent, S., Pel, M.J.C., Tommassen, J., Pieterse, C.M.J., van Kessel, K.P.M., and van Strijp, J.A.G. (2011). Pseudomonas evades immune recognition of flagellin in both mammals and plants. PLoS pathogens *7*, e1002206-e1002206.

Bauer, Z., Gomez-Gomez, L., Boller, T., and Felix, G. (2001). Sensitivity of different ecotypes and mutants of Arabidopsis thaliana toward the bacterial elicitor flagellin correlates with the presence of receptor- binding sites. The Journal of biological chemistry *276*, 45669-45676.

Belkhadir, Y., Yang, L., Hetzel, J., Dangl, J.L., and Chory, J. (2014). The growth-defense pivot: crisis management in plants mediated by LRR-RK surface receptors. Trends Biochem Sci *39*, 447-456.

Beyer, M., Nesterov, A., Block, I., Konig, K., Felgenhauer, T., Fernandez, S., Leibe, K., Torralba, G., Hausmann, M., Trunk, U.*, et al.* (2007). Combinatorial synthesis of peptide arrays onto a microchip. Science *318*, 1888.

Bodenhausen, N., Horton, M.W., and Bergelson, J. (2013). Bacterial communities associated with the leaves and the roots of Arabidopsis thaliana. PLoS One *8*, e56329.

Boutrot, F., and Zipfel, C. (2017). Function, Discovery, and Exploitation of Plant Pattern Recognition Receptors for Broad-Spectrum Disease Resistance. Annual Review of Phytopathology *55*, 257-286.

Buscaill, P., Chandrasekar, B., Sanguankiattichai, N., Kourelis, J., Kaschani, F., Thomas, E.L., Morimoto, K., Kaiser, M., Preston, G.M., Ichinose, Y.*, et al.* (2019). Glycosidase and glycan polymorphism control hydrolytic release of immunogenic flagellin peptides. Science *364*.

Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics *25*, 1972-1973.

Cesbron, S., Paulin, J.-P., Tharaud, M., Barny, M.-A., and Brisset, M.-N. (2006). The alternative σ factor HrpL negatively modulates the flagellar system in the phytopathogenic bacterium Erwinia amylovora under hrp-inducing conditions. FEMS Microbiology Letters *257*, 221-227.

Chen, P., and Zhang, J. (2020). Antagonistic pleiotropy conceals molecular adaptations in changing environments. Nature Ecology & Evolution *4*, 461-469.

Chen, T., Nomura, K., Wang, X., Sohrabi, R., Xu, J., Yao, L., Paasch, B.C., Ma, L., Kremer, J., Cheng, Y.*, et al.* (2020). A plant genetic network for preventing dysbiosis in the phyllosphere. Nature *580*, 653-657.

Chesnokova, O., Coutinho, J.B., Khan, I.H., Mikhail, M.S., and Kado, C.I. (1997). Characterization of flagella genes of Agrobacterium tumefaciens, and the effect of a bald strain on virulence. Mol Microbiol *23*, 579-590.

Chinchilla, D., Bauer, Z., Regenass, M., Boller, T., and Felix, G. (2006). The Arabidopsis receptor kinase FLS2 binds flg22 and determines the specificity of flagellin perception. The Plant cell *18*, 465-476.

Chinchilla, D., Zipfel, C., Robatzek, S., Kemmerling, B., Nurnberger, T., Jones, J.D., Felix, G., and Boller,
T. (2007). A flagellin-induced complex of the receptor FLS2 and BAK1 initiates plant defence. Nature *448*, 497-500.

Clarke, C.R., Chinchilla, D., Hind, S.R., Taguchi, F., Miki, R., Ichinose, Y., Martin, G.B., Leman, S., Felix, G., and Vinatzer, B.A. (2013). Allelic variation in two distinct Pseudomonas syringae flagellin epitopes modulates the strength of plant immune responses but not bacterial motility. The New phytologist *200*, 847- 860.

Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. Genome Res *14*, 1188-1190.

Cull, M.G., and Schatz, P.J. (2000). Biotinylation of proteins in vivo and in vitro using small peptide tags. Methods in enzymology *326*, 430-440.

Dangl, J.L., Horvath, D.M., and Staskawicz, B.J. (2013). Pivoting the plant immune system from dissection to deployment. Science *341*, 746-751.

Delmotte, N., Knief, C., Chaffron, S., Innerebner, G., Roschitzki, B., Schlapbach, R., von Mering, C., and Vorholt, J.A. (2009). Community proteogenomics reveals insights into the physiology of phyllosphere bacteria. Proceedings of the National Academy of Sciences *106*, 16428-16433.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high

throughput. Nucleic Acids Res *32*, 1792-1797.

Fairhead, M., and Howarth, M. (2015). Site-specific biotinylation of purified proteins using BirA. Methods in molecular biology (Clifton, NJ) *1266*, 171-184.

Felix, G., Duran, J.D., Volko, S., and Boller, T. (1999). Plants have a sensitive perception system for the most conserved domain of bacterial flagellin. The Plant journal : for cell and molecular biology *18*, 265- 276.

Fliegmann, J., and Felix, G. (2016). Immunity: Flagellin seen from all sides. Nature plants *2*, 16136.

Fürst, U., Zeng, Y., Albert, M., Witte, A.K., Fliegmann, J., and Felix, G. (2020). Perception of Agrobacterium tumefaciens flagellin by FLS2(XL) confers resistance to crown gall disease. Nature plants *6*, 22-27.

Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nature methods *6*, 343-345.

Gomez-Gomez, L., and Boller, T. (2000). FLS2: an LRR receptor-like kinase involved in the perception of the bacterial elicitor flagellin in Arabidopsis. Molecular cell *5*, 1003-1011.

Gouveia-Oliveira, R., and Pedersen, A.G. (2007). Finding coevolving amino acid residues using row and column weighting of mutual information and multi-dimensional amino acid representation. Algorithms for Molecular Biology *2*, 12.

Ha, D.G., Kuchma, S.L., and O'Toole, G.A. (2014). Plate-based assay for swarming motility in Pseudomonas aeruginosa. Methods in molecular biology (Clifton, NJ) *1149*, 67-72.

Hacquard, S., Spaepen, S., Garrido-Oter, R., and Schulze-Lefert, P. (2017). Interplay Between Innate Immunity and the Plant Microbiota. Annu Rev Phytopathol *55*, 565-589.

Hashimoto, Y., Zhang, S., and Blissard, G.W. (2010). Ao38, a new cell line from eggs of the black witch moth, Ascalapha odorata (Lepidoptera: Noctuidae), is permissive for AcMNPV infection and produces high levels of recombinant proteins. BMC biotechnology *10*, 50.

Hind, S.R., Strickler, S.R., Boyle, P.C., Dunham, D.M., Bao, Z., O'Doherty, I.M., Baccile, J.A., Hoki, J.S., Viox, E.G., Clarke, C.R.*, et al.* (2016). Tomato receptor FLAGELLIN-SENSING 3 binds flgII-28 and activates the plant immune system. Nature plants *2*, 16128.

Hmelo, L.R., Borlee, B.R., Almblad, H., Love, M.E., Randall, T.E., Tseng, B.S., Lin, C., Irie, Y., Storek, K.M., Yang, J.J.*, et al.* (2015). Precision-engineering the Pseudomonas aeruginosa genome with two-step allelic exchange. Nat Protoc *10*, 1820-1841.

Hohmann, U., Lau, K., and Hothorn, M. (2017). The Structural Basis of Ligand Perception and Signal Activation by Receptor Kinases. Annu Rev Plant Biol *68*, 109-137.

Hohmann, U., Santiago, J., Nicolet, J., Olsson, V., Spiga, F.M., Hothorn, L.A., Butenko, M.A., and Hothorn, M. (2018). Mechanistic basis for the activation of plant membrane receptor kinases by SERK- family coreceptors. Proceedings of the National Academy of Sciences of the United States of America *115*, 3488-3493.

Iwasaki, A., and Medzhitov, R. (2004). Toll-like receptor control of the adaptive immune responses. Nature immunology *5*, 987-995.

Iwasaki, A., and Medzhitov, R. (2015). Control of adaptive immunity by the innate immune system. Nature immunology *16*, 343-353.

Jaillais, Y., Belkhadir, Y., Balsemão-Pires, E., Dangl, J.L., and Chory, J. (2011). Extracellular leucine-rich repeats as a platform for receptor/coreceptor complex formation. Proceedings of the National Academy of Sciences *108*, 8503.

Kadota, Y., Sklenar, J., Derbyshire, P., Stransfeld, L., Asai, S., Ntoukakis, V., Jones, Jonathan D., Shirasu, K., Menke, F., Jones, A.*, et al.* (2014). Direct Regulation of the NADPH Oxidase RBOHD by the PRR- Associated Kinase BIK1 during Plant Immunity. Molecular cell *54*, 43-55.

Kappal, S. (2019). Data Normalization Using Median & Median Absolute Deviation (MMAD) based Z- Score for Robust Predictions vs. Min-Max Normalization.

Karasov, T.L., Almario, J., Friedemann, C., Ding, W., Giolai, M., Heavens, D., Kersten, S., Lundberg, D.S., Neumann, M., Regalado, J.*, et al.* (2018). Arabidopsis thaliana and Pseudomonas Pathogens Exhibit Stable Associations over Evolutionary Timescales. Cell Host Microbe *24*, 168-179.e164.

King, E.O., Ward, M.K., and Raney, D.E. (1954). Two simple media for the demonstration of pyocyanin and fluorescin. J Lab Clin Med *44*, 301-307.

Kovach, M.E., Elzer, P.H., Hill, D.S., Robertson, G.T., Farris, M.A., Roop, R.M., 2nd, and Peterson, K.M. (1995). Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. Gene *166*, 175-176.

Kozma, P., Hamori, A., Cottier, K., Kurunczi, S., and Horvath, R. (2009). Grating coupled interferometry for optical sensing. Applied Physics B *97*, 5-8.

Laflamme, B., Dillon, M.M., Martel, A., Almeida, R.N.D., Desveaux, D., and Guttman, D.S. (2020). The pan-genome effector-triggered immunity landscape of a host-pathogen interaction. Science *367*, 763-768.

Lampropoulos, A., Sutikovic, Z., Wenzl, C., Maegele, I., Lohmann, J.U., and Forner, J. (2013).

GreenGate--a novel, versatile, and efficient cloning system for plant transgenesis. PLoS One *8*, e83043.

Lee, H.S., and Belkhadir, Y. (2020). Damage Control: Cellular Logic in the Root Immune Response. Cell Host Microbe *27*, 308-310.

Leys, C., Ley, C., Klein, O., Bernard, P., and Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. Journal of Experimental Social Psychology *49*, 764–766.

Liberati, N.T., Urbach, J.M., Miyata, S., Lee, D.G., Drenkard, E., Wu, G., Villanueva, J., Wei, T., and Ausubel, F.M. (2006). An ordered, nonredundant library of Pseudomonas aeruginosa strain PA14 transposon insertion mutants. Proceedings of the National Academy of Sciences of the United States of America *103*, 2833-2838.

Lin, W., Li, B., Lu, D., Chen, S., Zhu, N., He, P., and Shan, L. (2014). Tyrosine phosphorylation of protein kinase complex BAK1/BIK1 mediates innate immunity. Proceedings of the National Academy of Sciences, 201318817.

Lozano-Duran, R., and Belkhadir, Y. (2017). A Technical Framework for Studying the Signaling Nexus of Brassinosteroids and Immunity. Methods in molecular biology (Clifton, NJ) *1564*, 49-61.

Lu, D., Wu, S., Gao, X., Zhang, Y., Shan, L., and He, P. (2010). A receptor-like cytoplasmic kinase, BIK1, associates with a flagellin receptor complex to initiate plant innate immunity. *107*, 496-501.

Macnab, R.M. (2003). How bacteria assemble flagella. Annu Rev Microbiol *57*, 77-100.

McCann, H.C., Nahal, H., Thakur, S., and Guttman, D.S. (2012). Identification of innate immunity elicitors using molecular signatures of natural selection. Proceedings of the National Academy of Sciences *109*, 4215-4220.

Meindl, T., Boller, T., and Felix, G. (2000). The bacterial elicitor flagellin activates its receptor in tomato cells according to the address-message concept. The Plant cell *12*, 1783-1794.

Millet, Y.A., Danna, C.H., Clay, N.K., Songnuan, W., Simon, M.D., Werck-Reichhart, D., and Ausubel,
F.M. (2010). Innate immune responses activated in Arabidopsis roots by microbe-associated molecular patterns. The Plant cell *22*, 973-990.

Mott, G.A., Thakur, S., Smakowska, E., Wang, P.W., Belkhadir, Y., Desveaux, D., and Guttman, D.S. (2016). Genomic screens identify a new phytobacterial microbe-associated molecular pattern and the cognate Arabidopsis receptor-like kinase that mediates its immune

elicitation. Genome Biol *17*, 98.

Naito, K., Taguchi, F., Suzuki, T., Inagaki, Y., Toyoda, K., Shiraishi, T., and Ichinose, Y. (2008). Amino acid sequence of bacterial microbe-associated molecular pattern flg22 is required for virulence. Molecular plant-microbe interactions : MPMI *21*, 1165-1174.

Nekrasov, V., Li, J., Batoux, M., Roux, M., Chu, Z.-H., Lacombe, S., Rougon, A., Bittel, P., Kiss-Papp, M., Chinchilla, D.*, et al.* (2009). Control of the pattern-recognition receptor EFR by an ER protein complex in plant immunity. The EMBO Journal *28*, 3428-3438.

Pel, M.J.C., van Dijken, A.J.H., Bardoel, B.W., Seidl, M.F., van der Ent, S., van Strijp, J.A.G., and Pieterse,
C.M.J. (2014). Pseudomonas syringae evades host immunity by degrading flagellin monomers with alkaline protease AprA. Molecular plant-microbe interactions : MPMI *27*, 603-610.

Poncini, L., Wyrsch, I., Dénervaud Tendon, V., Vorley, T., Boller, T., Geldner, N., Métraux, J.-P., and Lehmann, S. (2017). In roots of Arabidopsis thaliana, the damage-associated molecular pattern AtPep1 is a stronger elicitor of immune signalling than flg22 or the chitin heptamer. PLOS ONE *12*, e0185808.

Ramos, H.C., Rumbo, M., and Sirard, J.C. (2004). Bacterial flagellins: mediators of pathogenicity and host immune responses in mucosa. Trends Microbiol *12*, 509-517.

Rietsch, A., Vallet-Gely, I., Dove, S.L., and Mekalanos, J.J. (2005). ExsE, a secreted regulator of type III secretion genes in Pseudomonas aeruginosa. Proceedings of the National Academy of Sciences of the United States of America *102*, 8006-8011.

Robatzek, S., Bittel, P., Chinchilla, D., Kochner, P., Felix, G., Shiu, S.H., and Boller, T. (2007). Molecular identification and characterization of the tomato flagellin receptor LeFLS2, an orthologue of Arabidopsis FLS2 exhibiting characteristically different perception specificities. Plant molecular biology *64*, 539-547.

Rossez, Y., Wolfson, E.B., Holmes, A., Gally, D.L., and Holden, N.J. (2015). Bacterial flagella: twist and stick, or dodge across the kingdoms. PLoS Pathog *11*, e1004483.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular cloning: a laboratory manual (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B.*, et al.* (2012). Fiji: an open-source platform for biological-image analysis. Nature methods *9*, 676-682.

Schulze, B., Mentzel, T., Jehle, A.K., Mueller, K., Beeler, S., Boller, T., Felix, G., and Chinchilla, D. (2010). Rapid heteromerization and phosphorylation of ligand-activated plant transmembrane receptors and their associated kinase BAK1. The Journal of biological chemistry *285*, 9444-9451.

Sitaraman, R. (2015). Pseudomonas spp. as models for plant-microbe interactions. Front Plant Sci *6*, 787- 787.
Smakowska, E., Kong, J., Busch, W., and Belkhadir, Y. (2016). Organ-specific regulation of growth- defense tradeoffs by plants. Curr Opin Plant Biol *29*, 129-137.

Smakowska-Luzan, E., Mott, G.A., Parys, K., Stegmann, M., Howton, T.C., Layeghifard, M., Neuhold, J., Lehner, A., Kong, J., Grunwald, K.*, et al.* (2018). An extracellular network of Arabidopsis leucine-rich repeat receptor kinases. Nature *553*, 342-346.

Stadler, V., Felgenhauer, T., Beyer, M., Fernandez, S., Leibe, K., Guttler, S., Groning, M., Konig, K., Torralba, G., Hausmann, M.*, et al.* (2008). Combinatorial synthesis of peptide arrays with a laser printer. Angewandte Chemie (International ed in English) *47*, 7132-7135.

Sun, W., Dunning, F.M., Pfund, C., Weingarten, R., and Bent, A.F. (2006). Within-species flagellin polymorphism in Xanthomonas campestris pv campestris and its impact on elicitation of Arabidopsis FLAGELLIN SENSING2-dependent defenses. The Plant cell *18*, 764-779.

Sun, Y., Li, L., Macho, A.P., Han, Z., Hu, Z., Zipfel, C., Zhou, J.M., and Chai, J. (2013). Structural basis for flg22-induced activation of the Arabidopsis FLS2-BAK1 immune complex. Science *342*, 624-628.

Taguchi, F., Yamamoto, M., Ohnishi-Kameyama, M., Iwaki, M., Yoshida, M., Ishii, T., Konishi, T., and Ichinose, Y. (2010). Defects in flagellin glycosylation affect the virulence of Pseudomonas syringae pv. tabaci 6605. Microbiology (Reading, England) *156*, 72-80.

Takai, R., Isogai, A., Takayama, S., and Che, F.S. (2008). Analysis of flagellin perception mediated by flg22 receptor OsFLS2 in rice. Molecular plant-microbe interactions : MPMI *21*, 1635-1642.

Teixeira, P.J.P.L., Colaianni, N.R., Fitzpatrick, C.R., and Dangl, J.L. (2019). Beyond pathogens: microbiota interactions with the plant immune system. Current Opinion in Microbiology *49*, 7-17.

Trdá, L., Fernandez, O., Boutrot, F., Héloir, M.-C., Kelloniemi, J., Daire, X., Adrian, M., Clément, C., Zipfel, C., Dorey, S.*, et al.* (2014). The grapevine flagellin receptor VvFLS2 differentially recognizes flagellin-derived epitopes from the endophytic growth-promoting bacterium Burkholderia phytofirmans and plant pathogenic bacteria. The New phytologist *201*, 1371-1384.

Venkataram, S., Monasky, R., Sikaroodi, S.H., Kryazhimskiy, S., and Kacar, B. (2020). Evolutionary stalling and a limit on the power of natural selection to improve a cellular module. Proceedings of the National Academy of Sciences *117*, 18582.
Vetter, M., Karasov, T.L., and Bergelson, J. (2016). Differentiation between MAMP Triggered Defenses in Arabidopsis thaliana. PLOS Genetics *12*, e1006068.

Wang, F., Burrage, A.M., Postel, S., Clark, R.E., Orlova, A., Sundberg, E.J., Kearns, D.B., and Egelman,
E.H. (2017). A structural model of flagellar filament switching across multiple bacterial species. Nature communications *8*, 960.

Wang, J., Grubb, L.E., Wang, J., Liang, X., Li, L., Gao, C., Ma, M., Feng, F., Li, M., Li, L., *et al.* (2018). A Regulatory Module Controlling Homeostasis of a Plant Immune Kinase. Molecular cell *69*, 493- 504.e496.

Wang, S., Sun, Z., Wang, H., Liu, L., Lu, F., Yang, J., Zhang, M., Zhang, S., Guo, Z., Bent, A.F., *et al.* (2015). Rice OsFLS2-Mediated Perception of Bacterial Flagellins Is Evaded by Xanthomonas oryzae pvs. oryzae and oryzicola. Mol Plant *8*, 1024-1037.

Wei, Y., Balaceanu, A., Rufian, J.S., Segonzac, C., Zhao, A., Morcillo, R.J.L., and Macho, A.P. (2020). An immune receptor complex evolved in soybean to perceive a polymorphic bacterial flagellin. Nature communications *11*, 3763.

Winsor, G.L., Griffiths, E.J., Lo, R., Dhillon, B.K., Shay, J.A., and Brinkman, F.S. (2016). Enhanced annotations and features for comparing thousands of Pseudomonas genomes in the Pseudomonas genome database. Nucleic Acids Res *44*, D646-653.

Wyrsch, I., Dominguez-Ferreras, A., Geldner, N., and Boller, T. (2015). Tissue-specific FLAGELLIN- SENSING 2 (FLS2) expression in roots restores immune responses in Arabidopsis fls2 mutants. The New phytologist *206*, 774-784.

Yonekura, K., Maki-Yonekura, S., and Namba, K. (2003). Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy. Nature *424*, 643-650.

Yoon, S.-i., Kurnasov, O., Natarajan, V., Hong, M., Gudkov, A.V., Osterman, A.L., and Wilson, I.A. (2012). Structural Basis of TLR5-Flagellin Recognition and Signaling. *335*, 859-864.

Zhou, F., Emonet, A., Dénervaud Tendon, V., Marhavy, P., Wu, D., Lahaye, T., and Geldner, N. (2020). Co-incidence of Damage and Microbial Patterns Controls Localized Immune Responses in Roots. Cell *180*, 440-453.e418.

Zipfel, C., Robatzek, S., Navarro, L., Oakeley, E.J., Jones, J.D., Felix, G., and Boller, T. (2004). Bacterial disease resistance in Arabidopsis through flagellin perception. Nature *428*, 764-767.

CHAPTER 3: A COMPLEX IMMUNE RESPONSE TO FLAGELLIN EPITOPE VARIATION IN COMMENSAL COMMUNITIES[3]

Section 3.1: Introduction

Plants and animals are constantly surveilling their extracellular environment for the presence of external threats, environmental conditions, and cell-derived messages. One method of surveillance utilizes Pattern Recognition Receptors (PRRs) to monitor the extracellular space for ligands that reveal microbiome conditions (Belkhadir, 2014, Steinbrenner, 2020, Ronald and Beutler, 2010). During microbial colonization, diverse microbial derived protein fragments called Microbial Associated Molecular Patterns (MAMPs) act as ligands for PRRs (Ronald and Beutler, 2010, Boutrot and Zipfel, 2017, Zhou and Zhang, 2020). In plants, MAMP recognition leads to an immune response capable of halting microbial growth, termed MAMP-Triggered Immunity (MTI) (Li et al., 2005, Hacquard et al., 2017). The analysis of bacterial genomes from plant-associated isolates indicate that commensal bacteria produce MAMPs that can be detected by cognate plant immune receptors (Garrido-Oter et al., 2018). However, commensals also produce diverse MAMPs (Teixeira et al., 2019). Thus, furthering our understanding of MAMP variability and functional consequences will improve our understanding of how plant-associated

[3] This chapter previously appeared in Cell Host & Microbe. The original citation is as follows: **Nicholas R. Colaianni*,** Katarzyna Parys*, Ho-Seok Lee*, …, Corbin D. Jones, Youssef Belkhadir, Jeffery L. Dangl, **A complex immune response to flagellin epitope variation in commensal communities**, *Cell Host & Microbe*, 2021, ISSN 1931-3128, https://doi.org/10.1016/j.chom.2021.02.006. Note that * indicates that these authors contributed equally to the work.
My contributions to this work included statistical analyses, ROS burst and root antagonism assays, Co-IPs, creation of the bacterial databases used in this study, all computational work, figure creation, writing. I also intellectually led the experimental design and overall data interpretation.

microbial communities persist in the face of an immune response adapted to restrict microbial growth.

One of the most well studied MAMPs is a peptide derived from the abundant bacterial flagellum protein, FliC (Felix et al., 1999, Fliegmann and Felix, 2016, Boutrot and Zipfel, 2017). In the reference plant *Arabidopsis thaliana* (hereafter Arabidopsis), a 22 amino acid epitope of FliC, termed flg22 is required for bacterial motility, but is also sufficient to induce MTI (Naito et al., 2008, Parys et al., Chapter 2, Gómez-Gómez et al., 1999). Flg22 peptides can be generated by an unidentified protease that degrades deglycosylated FliC proteins (Buscaill et al., 2019). On the cell surface, the recognition of and response to flg22 in Arabidopsis is mediated by a two-receptor system. First, the flg22 peptide binds to FLAGELLIN SENSING 2 (FLS2) (Chinchilla et al., 2006). The C-terminal region of the flg22 peptide then induces the binding of FLS2 to a co-receptor BRI1-ASSOCIATED KINASE 1 (BAK1) (Chinchilla et al., 2007). This binding strategy defines the "address-message" concept of flg22 recognition: the 'address' is the N-terminal portion of the peptide that defines its ability to bind FLS2, while the 'message' is the C-terminal portion of the peptide that defines the formation of the FLS2-BAK1 complex. FLS2-BAK1 complex formation drives subsequent signaling that leads to functional MTI (Meindl et al., 2000, Sun et al., 2013).

The formation of the flg22-FLS2-BAK1 complex triggers a phosphorylation cascade that results in a set of temporally distinct responses. Within minutes, downstream signaling components are phosphorylated, including BOTRYTIS-INDUCED KINASE 1 (BIK1), which directly activates the production of a Reactive Oxygen Species (ROS) burst by phosphorylating the NADPH oxidase RbohD (Lu et al., 2009, Li et al. 2014, Kadota et al., 2014). Flg22 also promotes a MAP kinase cascade that leads to rapid transcriptional reprogramming, indicative of

MTI activation (Asai et al., 2002). Temporally later responses were identified from chronic exposure to immunogenic flg22. One of these chronic responses is seedling growth inhibition (SGI), illustrated by stunted seedling growth due to a growth-defense tradeoff (Gómez-Gómez et al., 1999, Belkhadir et al., 2011, Albrecht et al., 2012).

FLS2 responses induced by flg22 in Arabidopsis roots are not only cell type-specific, but also require cellular damage for maximal output (Zhou et al., 2020, Emonet et al., 2020, Rich-Griffin et al., 2020). These distinct flg22-induced responses indicate that FLS2 activities are adaptable and carefully deployed. This is supported by the specificity of immune elicitation across MAMPs (Vetter et al., 2016). To date, studies on the immune responses produced by a particular MAMP have mostly focused on a handful of epitope variants derived from pathogens and beneficial bacteria. Thus, there is a major gap in our understanding of the spectrum of responses engendered by natural MAMP variation present in commensal communities, and the breadth of responses to MAMPs, like those engendered by flg22, remains undefined.

Here, we leveraged advancements in the genomic characterization of hundreds of Arabidopsis commensal bacterial strains to address these knowledge gaps. We identified and screened, at previously unprecedented scales, the natural flg22 diversity present in commensal communities. First, we generated a community accessible database comprised of culturable root and leaf commensal bacteria all derived from Arabidopsis to query flg22 peptide diversity and signaling. Second, we screened >90 of these flg22 variants for the production of MTI responses. We found that commensal bacteria encode significant flg22 sequence diversity, and that a majority of this diversity results in flg22 peptides that evade defense activation. Amongst the evading flg22 variants identified in our primary screen, we identified flg22 peptides that antagonize the formation of the FLS2-BAK1 complex induced by active flg22 peptides. In the process of

characterizing FLS2 antagonists, we also identified flg22 variants that prime the early temporal responses to subsequent flg22 elicitation, but alter temporally late responses. We demonstrate that antagonistic flg22 peptides are found in prevalent commensal community members, and that a commensal strain functionally antagonizes the canonical flg22 response *in vivo*. We uncover flg22 variants that uncouple FLS2-BAK1-mediated signal transduction pathways. We also observe that evading variants are dominant in a large synthetic community experiment, but that immune-activating flg22 variants are enriched under an abiotic stress condition that represses the immune system. Thus, our data demonstrate that there is pervasive evasion of immune activation through reducing FLS2 activation and through the manipulation of immune outputs by receptor antagonism or modulation. This work provides a framework to better understand how commensal microbiomes assemble in the interface of immune surveillance.

Section 3.2: Results

3.2.1: Arabidopsis commensal bacteria encode substantial variation in the flg22 epitope of FliC

To identify the flg22 diversity present in natural communities, we built a database of 627 genomes derived from bacteria isolated from healthy Arabidopsis plants (Table S1; Bai et al., 2015, Levy et al., 2018). We identified 779 FliC proteins from 61% of the microbial genomes in the Arabidopsis database (Figure 3.1A). We created a phylogenetic tree using primary sequence similarity and classified the FliCs into 3 clades (Figure 3.1A; Figure 3.8A). We found that FliC clade assignment was indicative of taxonomic origin. Clade 1 FliCs are primarily found in Betaproteobacteria, Gammaproteobacteria, and Sphingomonadales; clade 2 FliCs are primarily found in Bacilli and Actinobacteria; and clade 3 FliCs are primarily found in Alphaproteobacteria, particularly Rhizobiales and Caulobacterales. The average primary

sequence identity of FliCs within clades 1, 2, and 3 was 47%, 58%, and 41%, respectively (Figure 3.1A). Bacterial motility has been demonstrated for bacteria that encode all three of these FliC clades (Kühn et al., 2018, Clarke et al., 2013, Haiko et al., 2013, Fuji et al., 2008, Iida et al., 2009), suggesting that adaptive levels of bacterial motility can be retained across the broad flagellin sequence divergence observed here.

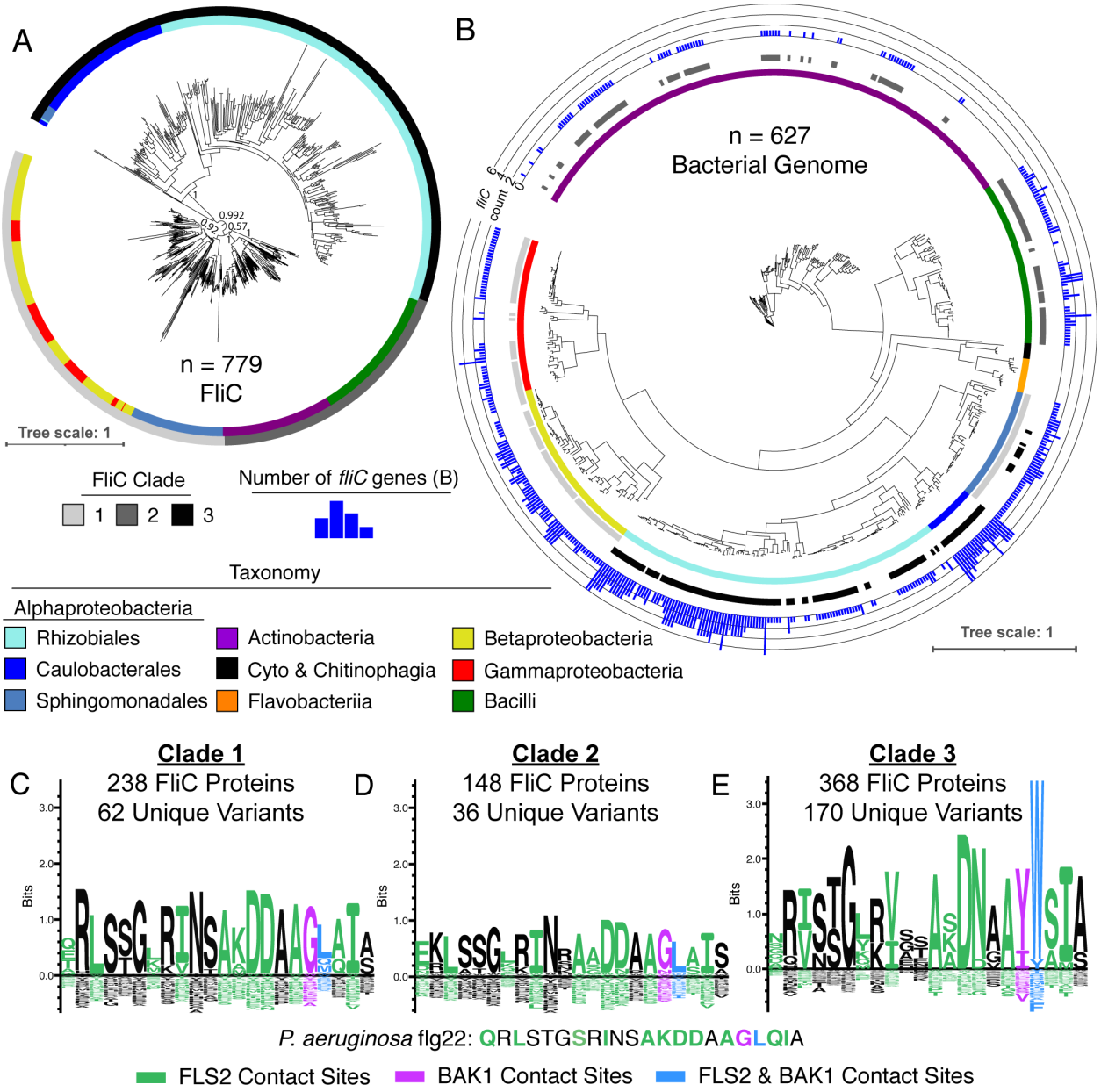


Figure 3.1. Taxonomically distinct FliC proteins contain three distinct clades of flg22 sequences. A) An illustrative tree produced using 779 high quality FliC proteins shows that the amino-acid sequence similarity of FliC proteins corresponds with bacterial taxonomy. From these

relationships we defined three clades based on sequence similarity and taxonomic relationships (More in Figure 3.8A). B) Bacterial genomes may contain multiple copies of *fliC* genes. A bacterial species tree of bacterial strains isolated from healthy Arabidopsis plants (Levy et al., 2018) that was created from a concatenated alignment of 31 single copy genes found in each microbe. The metadata from the inner to outer circles represent: Taxonomy, FliC clade any of their flagellin genes come from, and the number of *fliC* genes within each genome. C, D, E) Sequence logos representing the unique flg22 sequences found within clades 1, 2, and 3 respectively. Letters are colored based on contact sites inferred from the crystal structure of the flg22-FLS2-BAK1 complex (Sun et al., 2013). A previously studied immunogenic flg22 sequence from *Pseudomonas aeruginosa* (Pa22) is shown for comparison.

To investigate whether the diversity of FliC proteins is reflected in flg22 sequence variation, we extracted 268 unique flg22 sequences from the FliC proteins in the database. The flg22 sequences from clades 1 and 2 resemble the canonical, MTI active flg22 variant from *Pseudomonas aeruginosa* (Pa22) (Figure 3.1C-D). The clade 3 flg22 sequences were the most divergent from Pa22 (Figure 3.1E). Importantly, residues interacting with FLS2, Pa22 Asp^{15} and Pa22 Gly^{18}, and BAK1, Pa22 Leu^{19} in the Pa22-BAK1-FLS2 crystal structure are commonly altered in clade 3 flg22 peptides (Sun et al., 2013, Figure 3.1E). This suggests that the natural diversification of flg22 sequences in clade 3 FliC proteins occurs on residues that are critical for detection by the plant immune system. We posit that flg22 sequence diversity in clade 3 has functional consequences for immunogenicity. This proposition is consistent with results from a large-scale synthetic analysis of flg22 variation detailed in Parys et al. (Chapter 2).

Within the clade 3 FliCs, ~50% have a unique flg22 sequence. This is twice the amount of diversity found in clades 1 and 2, in which only ~25% FliC proteins have unique flg22 sequences. To understand how clade 3 FliCs became so diverse, we investigated *fliC* copy number (Figure 3.1B, Figure 3.8B). We found a median of two *fliC* genes in genomes with clade 3 FliCs, while only a median of one for clades 1 and 2 (Figure 3.1B). Up to six *fliC* gene products can be incorporated into the flagellum (Iida et al., 2009). Within genomes containing multiple *fliC* genes, we identified operons where *fliC* genes occur within 10 kB of one another,

suggesting that *fliC* expansion is the result of gene duplication (Figure 3.8B). Thus, diversification of the flg22 region in clade 3, especially the prominent changes of Asp^{15}, Gly^{18}, and Leu^{19}, may have been aided by *fliC* gene duplication, the maintenance of multiple *fliC* genes and antagonistic pleiotropy as detailed in the accompanying manuscript (Parys, et al., Chapter 2).

3.2.2: Pervasive evasion of FLS2 activation by commensal flg22 peptide variants

To characterize functionally the diverse flg22 variants found in Arabidopsis commensals, we synthesized 97 variants covering >30% of the flg22 diversity across all three FliC clades (N = Clade 1: 23, Clade 2: 17, Clade 3: 57) along with four control flg22 peptides (Table S2). We first screened the peptides for ROS burst (early response) and seedling growth inhibition (SGI, late response) induction on Arabidopsis plants.

We used the active *Pseudomonas syringae* pv. *tabaci* immunogenic flg22 peptide (Pta22) as a positive control and a variant with an Asp^{14} to Ala substitution (PtaDA), that is inactive at 10 nM but weakly active at 100 nM, as our baseline (Naito et al., 2008; Figures 3.9A-B). This allowed us to quantify ROS activity across a dynamic range between the highly active Pta22 peptide variant and the much less active mutant PtaDA peptide variant (Figure 3.9A-B). We also included the canonical immunogenic Pa22 variant (Figure 3.9A-B) and an inactive two amino acid truncated version of Pa22, hereafter called Pa20, as additional controls (Bauer et al., 2001). Strikingly, 64% (62/97) of the flg22 variants tested did not induce a ROS burst at either 10 nM or 100 nM (Figure 3.2A, Figure 3.9D). Of these, 57/62 peptides belong to clade 3, while the remaining five were split between clade 1 and clade 2 (four and one flg22 sequences, respectively). The 35/97 active flg22 variants produced ROS at both 10 nM and 100 nM (Figure 3.2A, Figure 3.9D). To ensure the flg22 variants were activating via FLS2, we assayed the active variants on Col-0 mutant plants that lack the FLS2 receptor (*fls2 efr,* see Methods). All 35 flg22

variants were inactive on *fls2 efr* mutant plants (Figure 3.9C). This demonstrates that FLS2 is capable of mediating a gradient of ROS burst responses from complete inactivity to that observed with 100nM of Pta22 or Pa22 (Figure 3.2A). Thus, in contrast to its singular all or none response to canonical epitopes, FLS2 allows a range of responses to low abundance immunogenic epitopes in a commensal landscape potentially dominated by non-immunogenic signals.

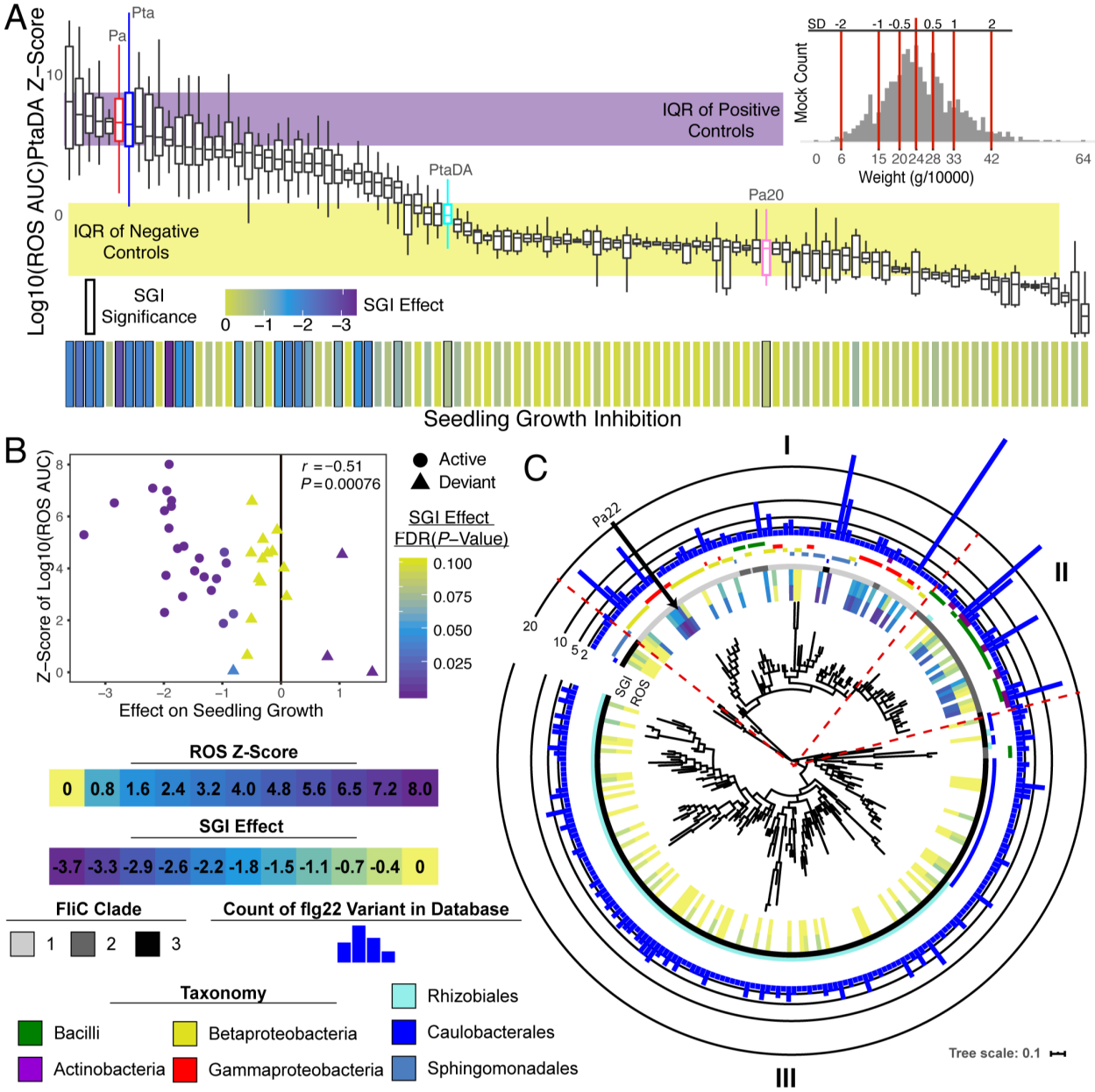


Figure 3.2. Evidence for pervasive evasion of flg22 induced MTI. A) Commensal bacteria contain both immunogenic and non-immunogenic flg22 variants. Boxplot: ROS burst profiles after the addition of 100 nM of the flg22 peptide of interest to wild type Col-0 leaf disks as measured by area under the curve (AUC) for at least 24 independent ROS burst profiles. Purple and yellow boxes indicate the combined interquartile range (IQR) of the positive (Pta22 and Pa22) and negative controls (PtaDA and Pa20), respectively. Barplot bottom: Plate-based Z-scores calculated for each flg22 peptide tested based on at least 16 independent SGI fresh weight measurements. Significance was calculated using. a linear-mixed model and is marked with black boxes based on a peptide having an FDR corrected *P*-value of less than 0.01. The color scale bar above indicates the estimated SGI effect. Top right histogram: The SGI effect is interpreted as the fresh weight differences between mock (median of 24 g/1000) and flg22 peptide treated plants normalized to the standard deviation (SD) of mock treated plants (~9 g/10000). B) ROS burst and SGI are significantly correlated, however there are exceptions where

ROS burst does not correspond to SGI. All peptides that induce ROS burst greater than PtaDA are represented by ROS (Z-score, Y-axis), SGI effect size (linear mixed model effects, X-axis) and *P*-values (SGI, color). The color scale corresponds to FDR corrected *P*-values calculated for the SGI linear mixed model effects. C) ROS burst and SGI discrepancies cluster based on flg22 sequence, FliC clade, and bacterial taxonomy. A phylogenetic tree from the amino acid sequences of 268 flg22 sequence variants. ROS burst (Z-score) and SGI (linear mixed model effects) shown are derived from A (inner two circles). The dashed red lines define boundaries for three groupings of flg22 peptide variants based on phenotypes, ROS+ and SGI+ (I), ROS+ and SGI- (II), and ROS- and SGI- (III). Count of flg22 variants represents the number of times that flg22 variant was found in the Arabidopsis database. Taxonomy is colored based on classification of the genome in which that flg22 sequence resides. FliC clade is colored based on the clade of FliC from which that flg22 variant derives.

We next performed SGI assays with 10 nM of each flg22 peptide variant. In these conditions, Pta22 and Pa22 readily reduced seedling fresh weight after 9 days of exposure (Figure 3.2A). Nearly all ROS inactive flg22 variants failed to induce significant SGI (Figure 3.2A). We found that 19/35 (54%) of our ROS active flg22 peptide variants induced significant SGI responses (Figure 3.2A) that were not observed on *fls2 efr* seedlings (Figure 3.9E). The remaining flg22 variants (16/35; 46%) were able to induce ROS but had no effect on SGI, and were termed 'deviant' peptide variants (Figure 3.2A-B). At least three classes of flg22 variants emerged: immunogenic peptides (20% of our tested variants), evading peptides that do not induce ROS burst or SGI (64% of our tested variants), and deviant peptides that only induce a ROS response (16% of our tested variants) (summarized in Figure 3.6).

Next, we constructed a phylogenetic tree using amino acid sequences from the flg22 variants to determine if ROS burst and SGI responses cluster by sequence (Figure 3.2C). We found that immunogenic flg22 sequences are derived mainly from clade 1 FliCs (I, Figure 3.2C). In contrast, non-immunogenic flg22 variants are mainly from clade 3 FliCs (III, Figure 3.2C) confirming our hypothesis that the high sequence divergence of flg22 epitopes in this clade has functional consequences on FLS2 responses. These data indicate that evasion of FLS2 activation is pervasive in Rhizobiales and Caulobacterales bacteria harboring clade 3 FliCs, potentially as a

requirement for optimal colonization of Arabidopsis. Interestingly, deviant flg22 variants are mainly derived from clade 2 FliC proteins (II, Figure 3.2C). Deviant flg22 variants share high similarity to Pa22, and are mainly found in Bacillus and Actinobacteria genomes (Figure 3.1D). Unexpectedly, the variation in these flg22 sequences occur in the 'address' portion of the peptide (Figure 3.1D), where a change of Gln^{1} to Glu and various changes at Lys^{13} are prominent in deviant peptides. Using synthetic flg22 variants, we found that $Pa22^{Q1E}$ increases ROS even with decreased binding affinity to FLS2 (Parys et al., Chapter 2). By contrast, the $Pa22^{K13D}$ leads to pronounced reductions in ROS levels by affecting the stable association of FLS2 to BAK1 (Parys et al., Chapter 2). We propose that the diversity in clade 2 flg22 peptides evolved to disrupt the sequential signaling outputs of FLS2, perhaps by a mechanism that involves the inappropriate recruitment of other regulators to FLS2. Overall, the clustering of flg22 responses based on taxonomic origin suggests that these responses play a role in community establishment and maintenance (Figure 3.2C). Our unprecedented screening of commensal-derived MAMP diversity demonstrates that a majority of commensal flg22 peptide variants evade FLS2 activation (64%) and do not significantly affect plant growth (80%).

3.2.3: Manipulation of immunity through receptor antagonism and signal modulation

Most flg22 variants from commensal bacteria evade FLS2 activation; but there are still many immunogenic variants. Microbial mechanisms to suppress MTI are beginning to be elucidated (Teixeira et al., 2019). One of these is receptor antagonism. Flg22 peptide variants that antagonize Pa22 induced phenotypes have been identified in *Ralstonia solanacearum, Pseudomonas cannabina* pv. *alisalensis* (ES4326), and among synthetic flg22 variants (Mueller et al., 2012, Clarke et al., 2013, Bauer et al., 2001). Parys et al. found that specific mutations in the C-terminal region of flg22 antagonized Pa22 by competing for FLS2 binding and blocking

the formation of a stable heterocomplex with BAK1 (Parys et al., Chapter 2). Considering the large number of flg22 peptide variants that evade FLS2 activation and carry variable C-terminal domains (Figures 3.1E, Figure 3.2C), we hypothesized that commensal bacteria have evolved flg22 peptide variants that block FLS2 activation.

To answer this, we tested flg22 peptides that failed to induce SGI for their ability to antagonize Pa22 induced expression of the root defense marker gene *Cytochrome P450 71A12 (CYP71A12)* (Millet et al., 2010). We found that 10 nM of Pa22 induced *proCYP71A12:GUS* expression in the root elongation zone, and the FLS2 receptor antagonist, Pa20, inhibited this response at a previously identified 10,000x ratio (100 μM) (Bauer et al., 2001; Figure 3.3A, Figure 3.10A). We tested 56 non-immunogenic flg22 variants and found 10 that consistently antagonized Pa22-triggered *proCYP71A12:GUS* expression at 100 μM (Figure 3.3A, Figure 3.10A). Interestingly, all antagonistic variants had mutations that could interfere with BAK1 binding at Gly^{18} and/or Leu^{19}, however two different N-terminal modification strategies emerged. One set increased the negative charge, while the other contained high sequence similarity to the Pa22 peptide (Figure 3.3L, Figure 3.10A). We hypothesize that both strategies support binding to FLS2 even with mutated C-terminal regions since they are able to block flg22 mediated *proCYP71A12:GUS* expression.

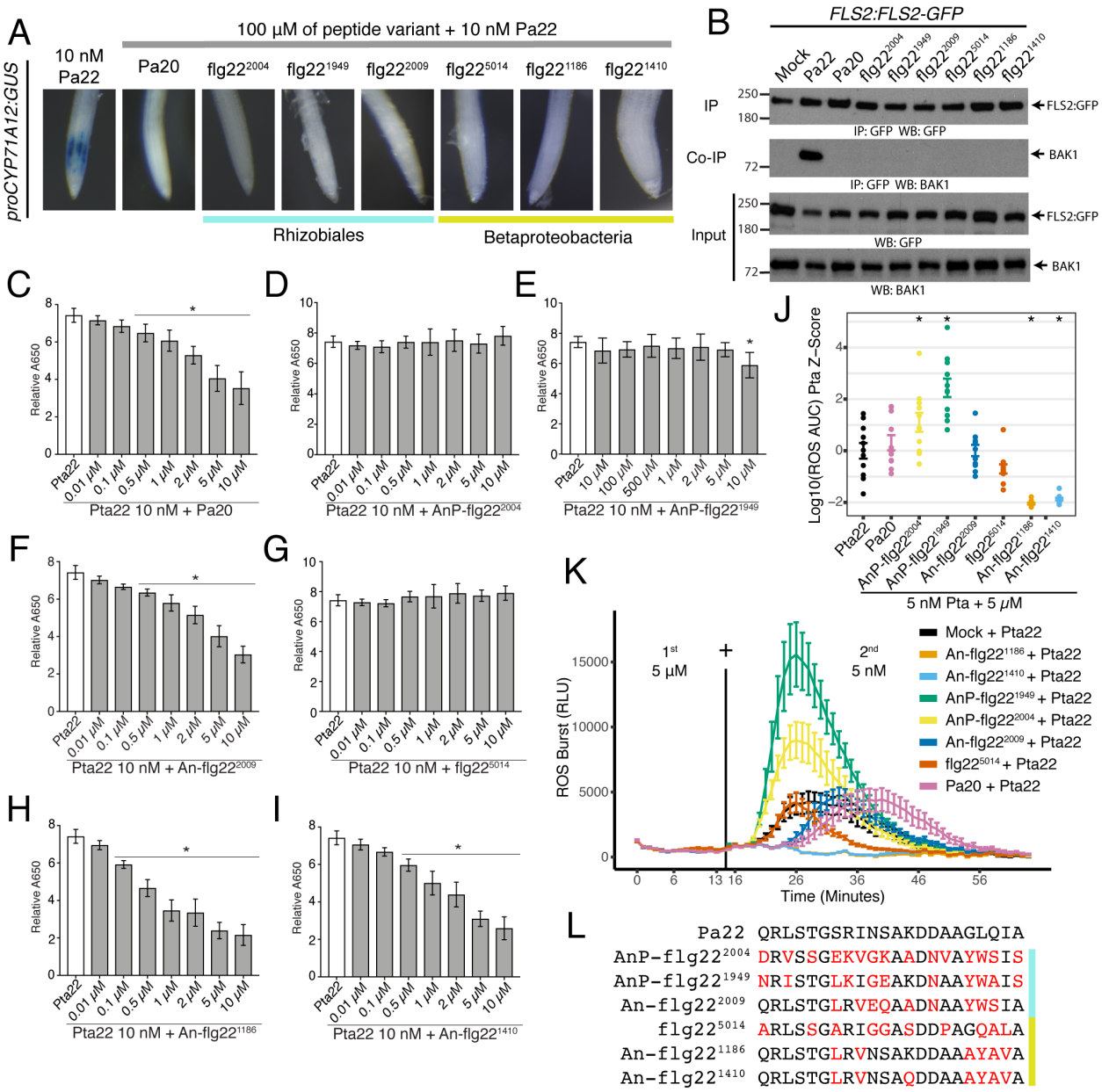


Figure 3.3. MTI altering flg22 variants are present in commensal Arabidopsis microbiomes. A) Some natural flg22 variants can antagonize Pa22 mediated *proCYP71A12:GUS* expression. 10 nM of Pa22 and 100 µM of each flg22 peptide were added to five to eight 7-day old *proCYP71A12:GUS* seedlings for five hours before GUS staining. Each peptide shown suppressed GUS expression in all roots in two or more independent experiments. Colors associated with each flg22 peptide ID indicate the taxonomic group the variant is derived from. Positive *proCYP71A12:GUS* expression is displayed by the 10 nM Pa22 condition. B) Potential commensal encoding antagonists do not induce FLS2-BAK1 complex formation. Co-IP using *proFLS2:FLS2-GFP* transgenic plants exposed to 1 µM of each flg22 variant for 15 minutes. Experiment was repeated three times with similar results. C-I) Flg22 peptides derived from commensal bacteria are able to antagonize FLS2-BAK1 heterocomplex formation measured *in vitro*. Shown is the relative absorbance (Abs 650 nm) over 2 h obtained in the presence of 10 nM Pta22 and at increasing concentrations of Pa20 and each potential antagonist. See figure 3.10B for more information on

the FLS2-BAK1 ectodomain interaction assay. Each barplot represents the mean and standard deviation from two independent biological experiments (n >= 5). Statistical significance was assessed using a one-way ANOVA followed by Dunnett's multiple comparison test (*= $P < 0.05$). J-K) Commensal flg22 peptides use different mechanisms to alter flg22 output responses in Arabidopsis. Flg22-induced ROS burst in wild type Col-0 plants driven by 5 nM of Pta22 after pre-exposing leaves to 5 μM of one of the antagonistic flg22 peptides indicated. Distilled water was used as mock treatment. The experiment was repeated three times with similar results; n=11. AnP = antagonistic primer; An = antagonist; see text for definitions. J) Integration of the area under the curve (AUC) measurements from ROS-burst assays on independent leaf disks. These values are Z-score normalized using the 5 nM of Pta22 and no pre-treatment condition. Bars represent the mean ± the standard error. Significant differences to the AUC of 5 nM Pta22 were determined using a one-way ANOVA followed by Dunnett's multiple comparison test (* = $P < 0.05$). K) The ROS-burst kinetics for the same AUC data shown in J. Each line represents the mean, and the bars represent the standard error of the measurements at each timepoint. 1st indicates the ROS burst induced by pre-exposure to the flg22 variants, and 2nd indicates the ROS induced after addition of the Pta22 peptide. L) Amino-acid sequence for each flg22 peptide analyzed. Differences to the Pa22 sequence are highlighted in red. Light blue and yellow indicates the flg22 variants are derived from Rhizobiales and Betaproteobacteria respectively.

To understand the molecular mechanism of antagonism for naturally derived flg22 peptides, we tested six variants that completely antagonized *proCYP71A12:GUS* expression and are representative of antagonist sequence diversity for their ability to induce FLS2-BAK1 complex formation in Arabidopsis (Figure 3.3A, Figure 3.3L). None of the six variants tested induced FLS2-BAK1 interaction in co-immunoprecipitation (Co-IP) assays (Figure 3.3B). This is consistent with these peptides' inability to elicit ROS and SGI.

Next, we tested if our candidate antagonist flg22 peptides were able to compete with Pta22 for the formation of FLS2-BAK1 heterocomplex *in vitro* (Figures 3.3C-I, Figure 3.10B). We found that three of the six peptide variants (An-flg22^{2009}, An-flg22^{1186} and An-flg22^{1410}) antagonized the FLS2-BAK1 interaction (Figures 3.3F, 3.3H, 3.3I, respectively). The An-flg22^{1186} peptide produced significant antagonistic activity at 10x the concentration of Pta22 (Figure 3.3H). To extend these results *in vivo*, we tested all six peptides for their ability to antagonize Pta22-mediated ROS burst after 15 minutes of pre-exposure (Figure 3.3J-K). We found that peptides An-flg22^{1186} and An-flg22^{1410} caused complete and persistent inhibition of

ROS burst, while An-flg22^{2009} behaved similar to Pa20 by delaying Pta22 mediated ROS burst (Figure 3.3J-K, Figure 3.10C). Interestingly, the other three antagonistic peptides either slightly reduced the timing of maximum ROS (flg22^{5014}, Figure 3.10C) or, via priming, increased the ROS burst produced by Pta22 (AnP-flg22^{1949}, AnP-flg22^{2004}; AnP means antagonistic primer, Figures 3.3J-K). These data demonstrate that receptor antagonists are encoded in commensal communities, and that some flg22 variants alter canonical MTI outputs in an incoherent manner. Thus, commensal organisms may have evolved specialized flg22 variants that exploit FLS2 signaling in order to facilitate plant colonization.

Motivated by the antagonistic oddities, we examined the ROS burst priming effect observed by the flg22 variants that antagonized Pa22-triggered *proCYP71A12:GUS* expression (Figure 3.3A), but did not inhibit Pta22-induced FLS2-BAK1 interaction (AnP-flg^{1949} and AnP-flg22^{2004}; Figures 3.3D-E). We found that Pta22-induced ROS burst scales with the concentration of AnP-flg22^{1949}, however the priming effect could not surpass the maximum response observed for Pta22 (Figures 3.10D-E). We then determined that pre-exposure of AnP-flg22^{1949} primes BIK1 phosphorylation (Figure 3.10F). We controlled for specificity, by testing AnP-flg22^{1949} priming activity towards another peptide MAMP, elf18 (Kunze et al., 2004), and found that AnP-flg22^{1949} does not prime the ROS burst elicited by elf18 (Figure 3.10G). These data indicate that antagonistic priming variants specifically modulate the response to immunogenic flg22, leading to the elimination of temporally late flg22 signaling outputs. Thus, our data define at least two ways to alter flg22 signaling output. The first is via interference with co-receptor recruitment and the second is by prevention of late output responses by an as yet unidentified mechanism that primes the canonical ROS response.

3.2.4: Active antagonistic peptides are encoded by prevalent community members

While commensals can produce antagonistic peptides, it is unclear if prevalent community members do so *in vivo*. Pathogenic *Ralstonia solanacearum* isolates were previously reported to encode a mix of antagonistic and evading flg22 variants (Mueller et al., 2012), prompting us to screen 9 flg22 variants found among 152 *Ralstonia* FliC proteins for antagonistic activity in Arabidopsis (Figure 3.11A; Table S4). Three of these, (*Ra* An-flg22^{1186}, *Ra* An-flg22^{1410}, *Ra* An-flg22^{5001}) were inactive in ROS burst and SGI assays and were antagonists of all immune outputs analyzed (*Ra* An-flg22^{1186}, *Ra* An-flg22^{1410} - Figures 3.3A-B, Figures 3.3H-K; *Ra* An-flg22^{5001} - Figures 3.4A-3.4C, Figures 3.11B). We identified 2 additional variants (*Ra* E-flg22^{5004} and *Ra* E-flg22^{5005}) that were inactive for ROS and SGI but did not act as antagonists, thus representing true evaders (Figures 3.4A-C, Figures 3.11B). Notably, we identified one variant (*Ra* flg22^{5003}) that acted as a weak antagonist (Figures 3.4A-C, Figures 3.11B).

We used a root enriched *Ralstonia* isolate (*Ra* CL21) that expresses the An-flg22^{1186} antagonist to test if *Ra* CL21 produces functional An-flg22^{1186} peptides (Levy et al., 2018). We replaced its native *fliC* with alleles encoding flg22 epitopes changed to *Ra* E-flg22^{5005} or *Ra* flg22^{5003}. CL21Δ*fliC*::*Ra* flg22^{5003} and CL21Δ*fliC*::*Ra* flg22^{5005} mutant strains retained 55% of *Ra* CL21 motility capacity (Figure 3.4D). *Ra* CL21 can suppress flg22-mediated root growth inhibition (RGI) and we reasoned that the antagonistic An-flg22^{1186} variant is at least partially responsible for this phenotype (Teixeira et al., submitted). We found that *Ra* CL21 can partially suppress RGI induced by 100nM of Pa22 in *UBQ10:FLS2-GFP* Arabidopsis plants that express FLS2 across all root cell types (Wyrsch et al., 2015) (Figure 3.4E). The *Ra* CL21 mutant expressing the weakly antagonistic *Ra* flg22^{5003} variant retained *Ra* CL21 levels of RGI

suppression (Figure 3.4E). Strikingly, we found that swapping *Ra* CL21 flg22 epitope to the *Ra* E-flg22^{5005} variant significantly reduced RGI suppression. This demonstrates that functional antagonists can be produced by prevalent commensal bacteria and can alter flg22 perception *in vivo*.

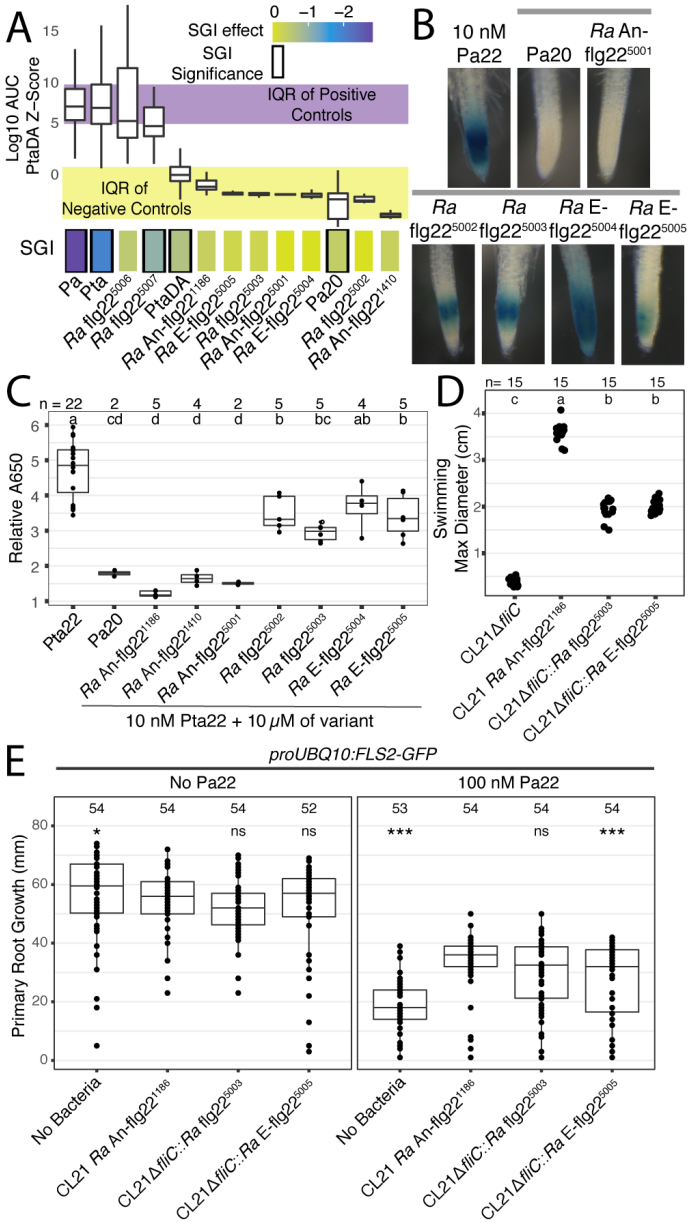


Figure 3.4. Prevalent commensal microbes produce flg22 antagonists. A) *Ralstonia* produce flg22 variants that evade immune elicitation. Boxplot: Area under the curve (AUC) was calculated for at least 24 independent ROS burst profiles after the addition of 100 nM of the flg22 peptide of interest to wild type Col-0 leaf disks. Purple and yellow boxes indicate the combined interquartile range (IQR) of the positive (Pta22 and Pa22) and negative controls (PtaDA and Pa20),

respectively. Barplot bottom: Plate-based Z-scores calculated for each flg22 peptide tested based on at least 16 independent SGI fresh weight measurements. Significance was assed using a linear mixed model and is marked with black boxes based on a peptide having an FDR corrected *P*-value of less than 0.01. The color scale bar on top shows the SGI effect compared to mock treated plants. In all panels E-flg22 indicates evading peptides and An-flg22 indicates antagonistic peptides that phenocopy those defined in Figure 3.3. B) *Ralstonia* flg22 variants that evade recognition can antagonize recognition of Pa22. Antagonism GUS assay where 10 nM of Pa22 and either 100 μM of Pa20 or a *Ralstonia* flg22 peptide were added to 7-day old *proCYP71A12:GUS* seedlings for five hours before GUS staining. Each peptide displayed as an antagonist suppressed GUS expression in all roots from at least two independent experiments. Positive *proCYP71A12:GUS* expression is displayed by the 10 nM Pa22 condition. All conditions with a grey bar indicate that 10 nM of Pa22. Was added with 100 μM of the variant. C) *Ralstonia* flg22 peptides quantitatively differ in antagonism of the FLS2-BAK1 heterocomplex formation measured *in vitro*. Shown is the relative absorbance (Abs 650 nm) over 2 h obtained in the presence of 10 nM Pta22 and 10 μM of Pa20 and each *Ralstonia* flg22 peptide. Statistically different groups displayed with letters at the top were identified using a one-way ANOVA followed by a Tukey's test ($\alpha = 0.05$). D) *Ralstonia* CL21 derivatives with mutant *fliC* genes containing swapped flg22 epitopes retain motility function. The CL21 strains indicated were inoculated in the center of 15 independent motility plates. The resulting swimming phenotype was scanned and the location with the largest diameter was measured using ImageJ. Statistically different groups were identified using a one-way ANOVA followed by a Tukey's test ($\alpha = 0.05$). E) CL21 strains produce antagonistic variants that suppress flg22-mediated root growth inhibition. The CL21 strains indicated were spread on plates supplemented with or without 100 nM of Pa22 at an OD_{600} of 0.0002. 7-day old *proUBQ10:FLS2-GFP* seedlings were transferred to these plates. Main root elongation was measured with a ruler seven days later. Main root elongation differences to WT CL21 within the Pa22 minus or plus conditions were evaluated with a linear mixed model (*= $P < 0.05$ and ***= $P < 0.001$) and numbers at the top indicate the number of roots measured.

We wondered whether antagonists were encoded in other prevalent commensal isolates, like the operational taxonomic unit (OTU) group defined in the phyllosphere by *Pseudomonas* strains *Pseudomonas OTU5* (Karasov et al., 2018). Genomes from this OTU encode 2 major flg22 epitopes, OTU5 An-flg22^{Pv1} and OTU5 flg22^{Pv2} found in 50% and 48% of the OTU5 genomes, respectively (Figure 3.11C; Karasov et al., 2018). The OTU5 An-flg22^{Pv1} variant has co-occurring mutations of Asn^{15} and Val^{18} that act as an antagonist in the Pa22 sequence background, and the OTU5 flg22^{Pv2} variant is similar to a previously identified atypical flg22 sequence encoded in *Pseudomonas cannabina* pv. *alisalensis* (ES4326) (Parys et al., Chapter 2, Clarke et al., 2013). Thus, we hypothesized that the major flg22 variants from this OTU group

would act as antagonists. We found that both OTU5 variants suppressed Pa22-mediated induction of *CYP71A12:GUS*, while only the OTU5 An-flg22^{Pv1} variant suppressed flg22 mediated ROS burst (Figures 3.11C-D). This is consistent with findings from Parys et al. (Chapter 2) demonstrating antagonism of the FLS2-BAK1 complex *in vivo.* These data indicate that intimately associated phyllosphere and root communities contain commensals that produce active antagonistic peptide variants, and that they may be produced in proportions that reduce immune system activities.

3.2.5 Natural flg22 variants drive separable MTI responses

Canonical flg22 peptides activate all known MTI responses. We defined deviant flg22 variants that induce ROS burst but not SGI (Figures 3.2A-B). To investigate their inability to activate full classical MTI defense markers, and to address whether the response to flg22 peptide variants is organ-specific, we monitored the activation of the transcriptional reporters *proMYB51:NLS-3mVenus* and *proWRKY11:NLS-3Venus* in Arabidopsis roots and epidermal cells of cotyledons, respectively (Poncini et al., 2017). ROS-inducing deviant peptide variants D-flg22^{1391}, D-flg22^{1857} and D-flg22^{1471} failed to induce the transcription of either reporter genes, as predicted from their inability to induce SGI (Figure 3.5A, Figures 3.12A-B). Unexpectedly, we identified a second deviant class, deviant-2, that induced transcriptional changes in our reporter lines similar to Pa22, but induced less ROS than PtaDA at 100 nM (Figure 3.5A, Figures 3.12A, Figure 3.9C; D2-flg22^{161}). This very low-level ROS burst was FLS2-dependent (Figure 3.9C). Previous work has reported the uncoupling of ROS burst and defense gene induction after MAMP recognition using chemical manipulation and mutant Arabidopsis lines (Lu et al., 2009, Shinya et al., 2014, Tran et al., 2020). However, to our knowledge, this is the first time that natural MAMP epitopes have been shown to decouple MTI outputs. This indicates that the

specificity of commensal-encoded flg22 peptide recognition by FLS2 may result in the recruitment of different signaling components, leading to heterogenous intracellular signaling outputs.

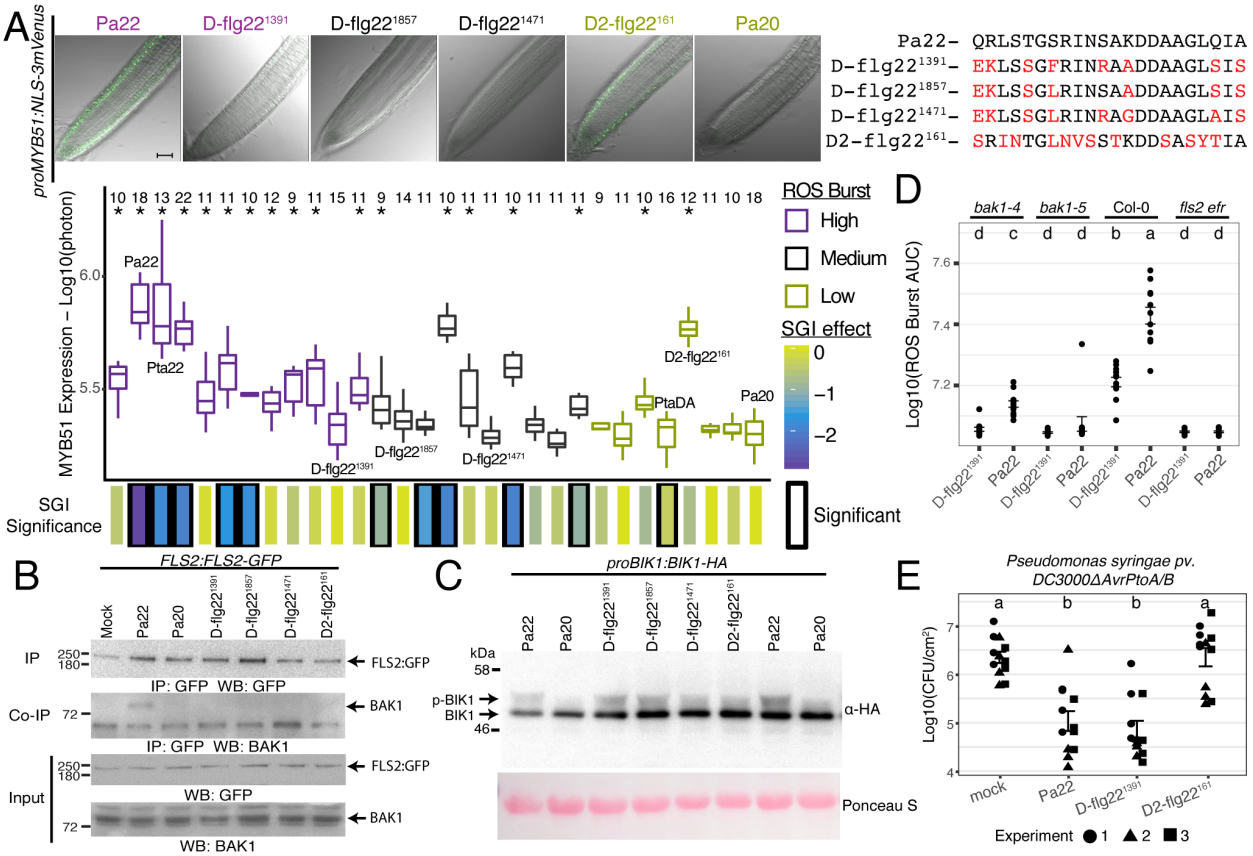


Figure 3.5. Deviant natural flg22 variants drive separable MTI responses. A) Flg22 variation can result in non-canonical gene expression changes. Top: Expression pattern of defense marker gene *MYB51* in Arabidopsis roots exposed to 10 nM of each four deviant (D; see text for definition) flg22 variants (numbered at top; color-coded by ROS burst activity). Shown are merged fluorescent and brightfield confocal microscope images of representative roots. The black scale bar represents 50 µm. Bottom: Fluorescence signal quantifications for microscopic observations of *proMYB51:NLS-3mVenus* expression after 10 nM exposure to each flg22 peptide for 24 hours. Peptides are ordered and color-coded by ROS burst activity (Figure 3.2A). The number above each boxplot represents the number of roots imaged for each variant. Significant *proMYB51:NLS-3mVenus* expression was determined by comparing each variant to the gene expression observed for Pa20 using an ANOVA and Dunnet's test (* = $P < 0.05$). The amino-acid sequences of deviant and deviant-2 peptides are shown to the right with differences to the Pa22 sequence highlighted in red. B) The deviant and deviant-2 peptides do not induce FLS2-BAK1 complex. Co-IP using *proFLS2:FLS2-GFP* transgenic plants exposed to 100 nM of each numbered flg22 peptide for 15 minutes. Western blot analyses of lysates were performed using anti-BAK1 and anti-GFP antibodies. This experiment was repeated three times with similar results. C) Deviant and deviant-2 peptides activate BIK1 phosphorylation. Western blot analyses of BIK1 phosphorylation in

proBIK1:BIK1-HA seedlings exposed to 100 nM of each peptide variant for 40 minutes. The experiment was repeated two times with similar results. D) ROS burst for D-flg22^{1391} is BAK1-dependent even though D-flg22^{1391} does not induce Co-IP between FLS2 and BAK1. Flg22-induced ROS burst of wild type Col-0, *fls2 efr*, *bak1*-5, and *bak1*-4 leaf disks exposed to D-flg22^{1391} or Pa22 at 100 nM. Shown are AUC measurements, where each error bar represents the mean ± the standard error of 12 independent leaf disks at each time point measured. Statistically different groups indicated by letters at the top were identified with a one-way ANOVA and Tukey's test ($\alpha = 0.05$). Experiment was repeated three times with similar results. E) The D-flg22^{1391} variant activates effective MTI, while D2-flg22^{161} does not. Wild type Col-0 plants were hand infiltrated with 100 nM of peptide or distilled water (mock) 24 hours before hand infiltration of *Pseudomonas syringae* pv. *tomato* DC3000Δ*AvrPtoA/B* at an OD_{600} of 0.0002. 12 data points from three independent experiments are displayed. Significantly different groups indicated at the top were identified using a two-way ANOVA controlling for batch effect and a Tukey test (α=0.05). The error bars indicate the mean ± standard error.

To identify the machinery necessary for the signaling of D-flg22^{1391}, D-flg22^{1857}, and D-flg22^{1471}, we tested if they induced FLS2-BAK1 complex formation *in vivo* with 100 nM of each peptide variant, and found that none did (Figure 3.5B). Strikingly however, these flg22 variants retained the ability to induce BIK1 phosphorylation (Figure 3.5C). We hypothesized that these variants likely activate an FLS2 receptor complex that does not include BAK1. We therefore tested the requirement of BAK1 for the ROS burst induced by deviant variants D-flg22^{1391}, D-flg22^{1857} and D-flg22^{1471} using two different BAK1 mutants, *bak1-4* (a null allele; Chinchilla et al., 2007) and *bak1-5* (a dominant negative allele; Schwessinger et al., 2011). We found that these three deviant flg22 variants required BAK1 for ROS burst production (Figure 3.5D, Figure 3.12C-D). This suggests that BAK1-FLS2 binding dynamics and/or complex-associated proteins are capable of fine-tuning flg22 signaling output.

To identify the functional significance of deviant and deviant-2 peptides in MTI, we performed flg22-driven protection assays. We hand inoculated leaves with 100 nM of Pa22, D-flg22^{1391}, or D2-flg22^{161}, 24 hours before hand inoculation with *Pseudomonas syringae* pv. *tomato* DC3000Δ*avrPtoA/B,* a weak pathogen lacking two Type III effector genes required to dampen MTI (He et al., 2006). We found that the Pa22 peptide reduced the colonization of

DC3000Δ*avrPtoA/B* significantly compared to plants inoculated with distilled water (Figure 3.5E). Interestingly, the deviant peptide variant, D-flg22^{1391}, limited the growth of DC3000Δ*avrPtoA/B* to the same level as Pa22, while the deviant-2 peptide variant, D2-flg22^{161}, did not reduce the growth of DC3000Δ*avrPtoA/B* (Figure 3.5E). This demonstrates that flg22 variants can induce distinct MTI responses that alter colonization outcomes. In sum, we define 6 classes of flg22 peptide variants based upon the immune outputs induced in mono-association or by how they affect the response to other flg22 peptides (Figure 3.6A).

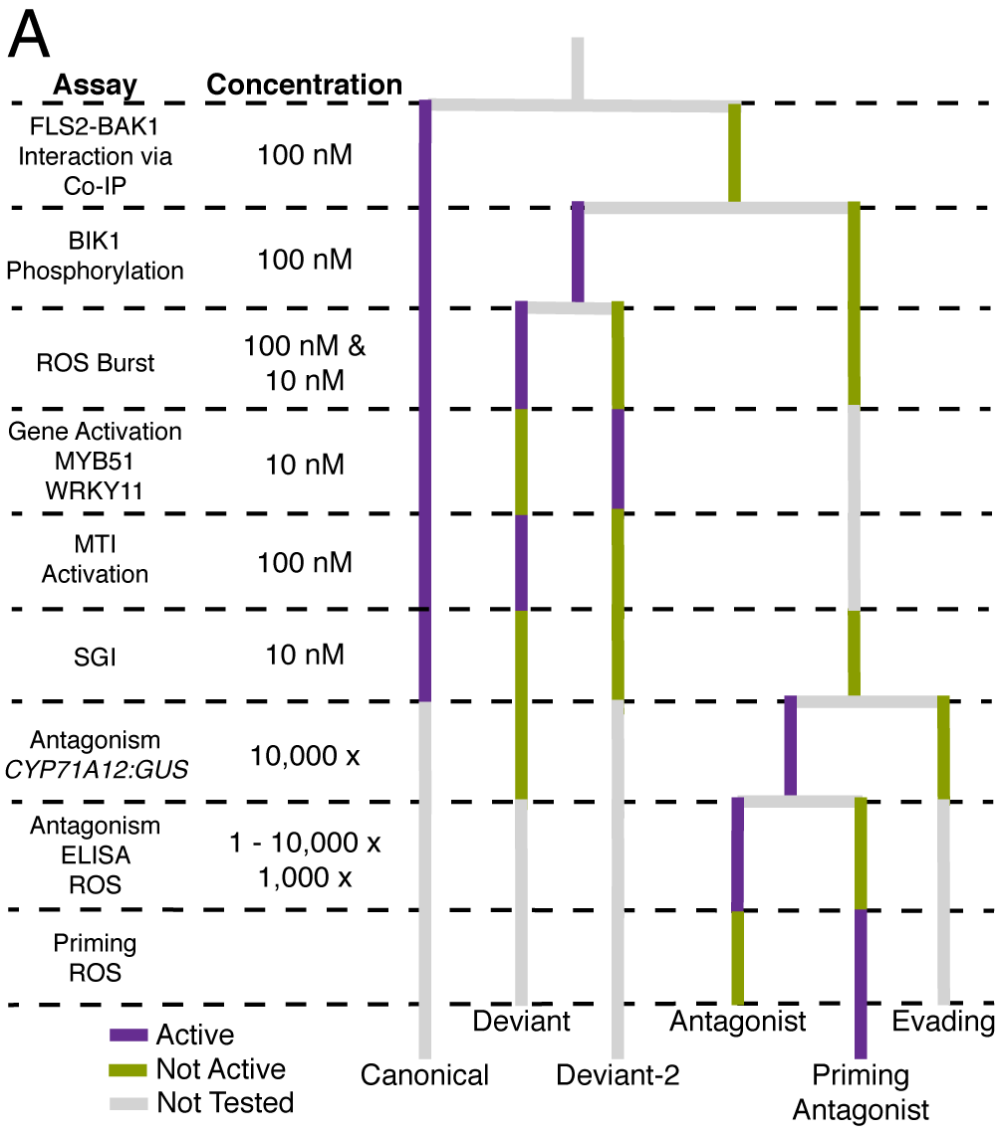


Figure 3.6. Commensal communities can produce at least 6 unique classes of flg22 variants, each of which results in a different response from the plant. Some of these may reshape the

interaction between the plant and other community members**.** A summary of our experiments that collectively defines 6 unique classes of flg22 variants.

3.2.6: Non-immunogenic variants dominate the flg22 functional repertoire in plant-associated communities

The composition of flg22 peptides in plant-associated communities may be actively shaped by plant immune surveillance. Alternatively, community homeostasis could be an epiphenomenal effect of flg22 variation on the immune system. In either case, the effects of flg22 variation on commensal community structure are unknown. We therefore leveraged existing synthetic community (SynCom) composition data from plant-associated microbiota assembled from a 185-member SynCom (Finkel et al., 2020) to analyze the flg22 functional repertoire (evading, antagonistic and deviant, as defined above). It is important to note that both our flg22 database and this 185-member community are based on culturable isolates and contain approximately 65% of the class level diversity found in natural communities (Figure 3.13A, Finkel et al., 2019). However, they nonetheless provide a good representation of the isolate fraction of plant-enriched communities (Figure 3.13A, Finkel et al., 2019). We accurately categorized all FliC variation and functionally characterized approximately two-thirds of the total flg22 abundance within the 185-SynCom communities. This included the relative abundance of flg22 variation and functional classes within and across plant-associated root and shoot fractions (Fig 3.13B). Importantly, we could almost completely annotate the relative abundance of flg22 variants from clades 1 and 2, which contain the active flg22 variants identified in our functional screen (Figure 3.13C, Figure 3.2C).

We first analyzed the FliC clade relative abundance in microbiota derived from the agar, root, and shoot fractions of seedlings colonized by SynCom185. We found a significant enrichment in bacteria encoding clade 3 FliCs in the root- and shoot-associated communities,

while bacteria encoding clade 1 FliCs were depleted (Figure 3.7A). Bacteria containing no *fliC* genes were unchanged across the fractions, suggesting that FliC type affects colonization (Figure 3.7A). To understand if the colonization effect of FliC clades varies with flg22 activity, we assigned the relative abundance of each unique V3-V4 16S sequence variant (Useq) derived from SynCom185 to all unique flg22 variants represented by the Useq. We then aggregated flg22 relative abundance values across flg22 functional classes. We found that evading peptide variants were enriched while immune active flg22 variants were depleted in plant-associated communities (Figure 3.7B). Sequence logos representing the flg22 functional capacity in root and shoot communities, resemble evasive flg22 variants based on the mutated C-terminal regions (Figure 3.7C-D). This data suggests evading flg22 variants are dominant and important in plant associated communities.

To gain insight into whether plants responses to abiotic stress alters flg22 composition, and hence community composition, we compared the communities assembled under optimal growth conditions to communities assembled under salt stress. Salt stress reduces plant defense through abscisic acid-mediated immune suppression (Berens et al., 2019). We hypothesized that communities formed under high salt would not be depleted in active variants or enriched in evading variants. We confirmed this (Figure 3.13D). Additionally, bacteria harboring the most immunogenic variants were now enriched in plant-associated communities (Figure 3.7E, Figure 3.13E). Thus, commensal MAMPs are likely to be actively monitored by plants and, as a result, flg22 variants that can evade may be selected for. This interplay among immunogenic variants, evading variants, antagonizing variants and the host immune system may ultimately promote community diversity and maintain microbial homeostasis.

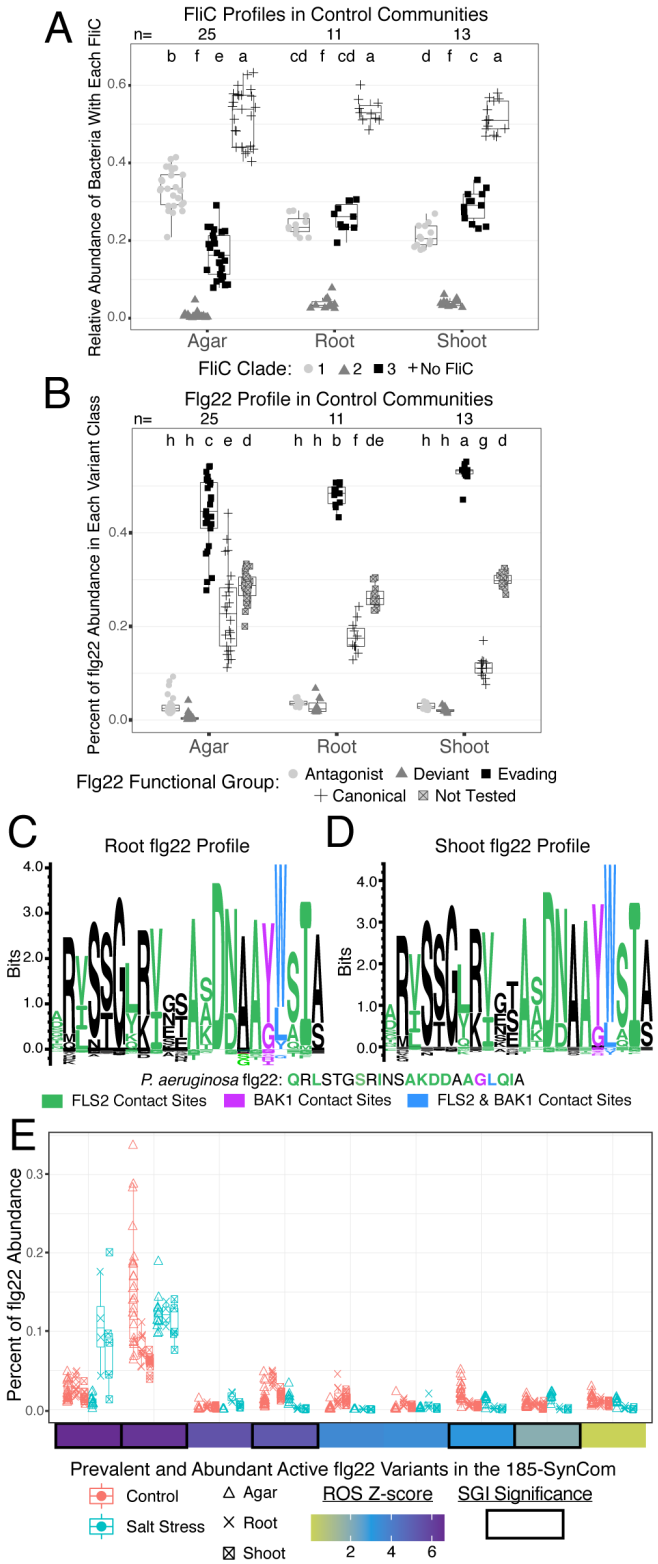


Figure 3.7. Non-immunogenic variants dominate the flg22 functional repertoire in plant-associated synthetic communities (SynCom). A) Plant-associated communities are enriched for bacteria encoding clade 3 FliCs, while being depleted in clade 1 FliCs. A 185-member SynCom was applied to Arabidopsis seedlings in normal growth conditions (21°C and 1000 μM Pi) and

were sequenced for the V3-V4 16S region (Finkel et al., 2020). The relative abundance values for each unique V3-V4 16S sequence variant in SynCom185 (Useq) were assigned to the FliC clade encoded in the various genomes. Statistically significant groups were identified using a two-way ANOVA and Tukey test (α=0.05) and are indicated by the letters at the top. B) The flg22 functional repertoire in plant-associated communities is dominated by evading peptide variants, while immunogenic flg22 variants are depleted. Useq relative abundance values were assigned to the flg22 variants represented by the Useq. Percentages were then calculated using the total abundance over all flg22 variants defined in the community. Statistically significant groups were identified using a two-way ANOVA and Tukey test (α=0.05) and are indicated by the letters at the top. C-D) The flg22 composition in root and shoot communities resembles immune evading flg22 variants. Sequence logos created from representing each flg22 variant based on its median relative abundance in root (C) and shoot (D) communities. For comparison, the Pa22 amino-acid sequence variant is shown. The colors represent contact sites inferred from the crystal structure of the Pa22-FLS2-BAK1 complex (Sun et al., 2013). E) Plant-associated communities assembled under salt stress are enriched in the most immunogenic flg22 variants. Prevalent and abundant flg22 sequences are defined as sequences having a flg22 percent abundance statistically different from 0. ROS (Z-score) and SGI significance are taken from Figure 3.2A. Flg22 percent abundances are taken from the data displayed in Figure 3.7B and Figure 3.13C.

Section 3.3: Discussion

Triggering of plant immune receptors by MAMPs classically results in an immune response that facilitates halting the growth of an invading microbe. Yet, plants stably associate with a compendium of microbes that can display collections of putative MAMPs closely related to those found expressed by pathogens. While mechanisms to suppress MTI are present in pathogens and beneficial microbes, it is unclear how common and effective they are in the commensal community context (Teixeira et al., 2019). Thus, to coexist with beneficial communities of microbes, without suffering the tissue damage associated with chronic immunity, plants must have evolved to unfold their response to MAMPs in ways that allow a balance between pathogen surveillance and commensal colonization. To address this, we focused on the immune responses engendered by flagellin epitope variation encoded in genomes of strains derived from commensal communities.

The primitive organisms *Hydra* and *Euprymna scolopes* utilize PRRs to recognize commensal bacteria present in their environment, suggesting that this may be an ancient function

of innate immunity (Chu and Mazmanian, 2013). We propose that FLS2 evolved to act as a communication device to allow plants and microbes to interact optimally. Thus, our findings challenge dogmas of a 'strict' immune signaling function for FLS2, and rather indicate that FLS2 functions as a community sensor that can be utilized to influence the assembly of specific commensal microbiota.

Commensal bacteria produce substantial flg22 diversity that contributes to pervasive evasion of MTI. We found that commensals contain abundant flg22 sequence diversity, and that most flg22 variants are non-immunogenic to the host. A majority of these flg22 peptides commonly introduce a mutation of Asp^{15} to Asn^{15}, which is not sufficient for immune evasion (Parys et al., Chapter 2). Rather, we attribute the 'evading' activities of these peptides to co-occurring polymorphisms in the 'message' domain of flg22. Collectively, our data support the notion that these variants evolved naturally to reduce their agonistic activity and prevent maximal FLS2-BAK1 interaction for full MTI activation. This is consistent with data using synthetic flg22 variants that identified evolutionary strategies by which *Pseudomonads* evolve evading flg22 variants in the face of motility loss (Parys et al., Chapter 2). As most of the evading flg22 variants come from bacterial genomes with multiple *fliC* genes, we propose that gene duplication facilitates the exploration of this evolutionary space by producing multiple functional gene copies (Ohno, 1979).

Flg22 response diversity may play a role in community composition and maintenance. We found that the responses induced by flg22 variants seem to reflect the colonization ability of the bacterial taxa in which they are found. Evasion of flg22 induced immunity is pervasive in Rhizobiales and Caulobacteriales isolates (Figure 3.2C), which are exemplary colonizers of Arabidopsis (Finkel et al., 2020, Carlström et al., 2019). Alternatively,

Gammaproteobacteria, which predominantly contain canonical flg22 variants (Figure 3.2C), are poor colonizers of Arabidopsis roots (Finkel et al., 2020, Carlström et al., 2019). We find that active variants are depleted, while evading variants are enriched in plant-associated communities. This is dependent on both flg22 epitope functional class and plant defense (Figures 3.7B and 3.13C). These results are not complete, as we only study flg22 variants and community dynamics from culturable isolates. However, these results support an active role of flg22 in community structuring. We speculate that this concept will translate to other MAMP signals and other host-microbiota interaction systems.

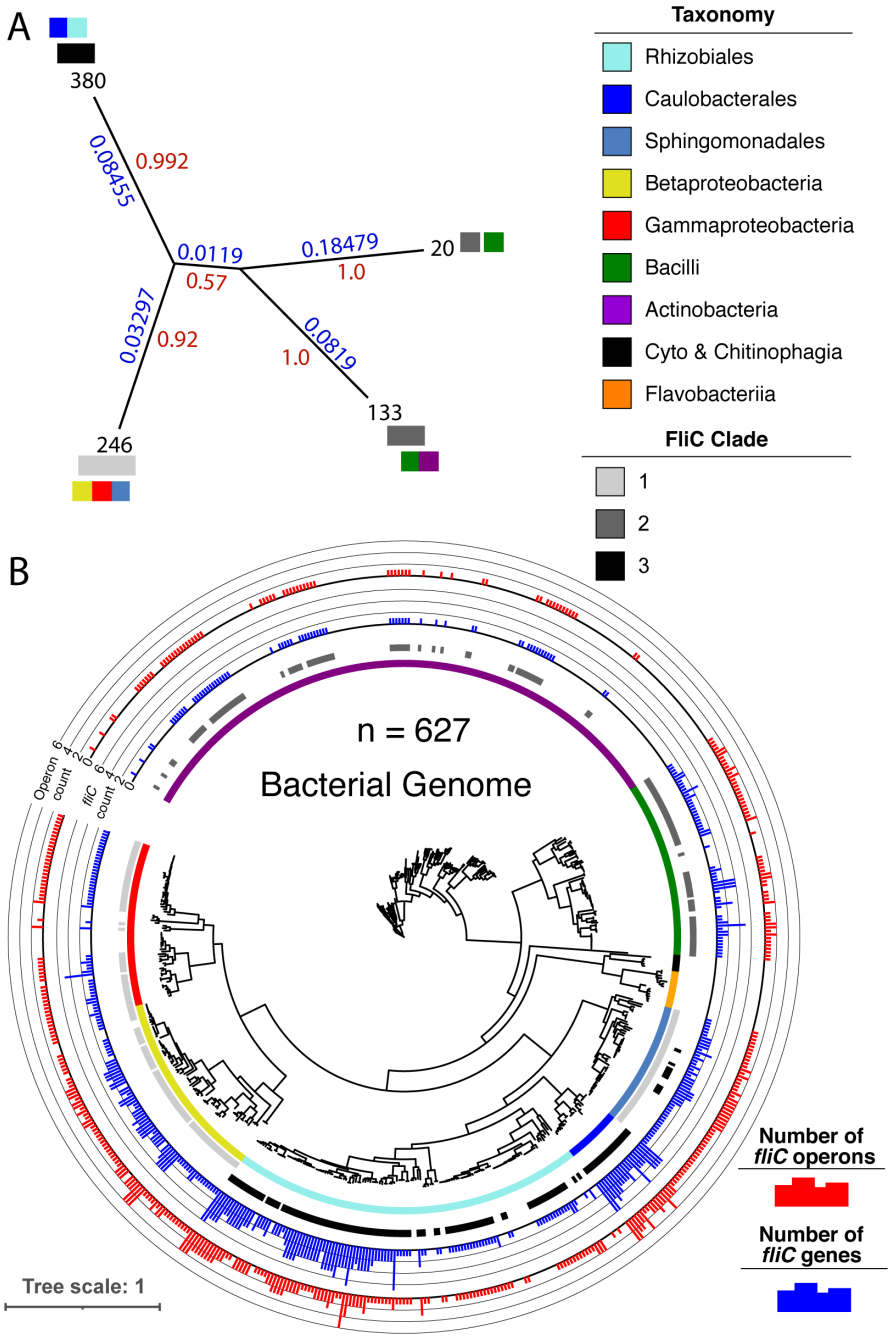
Modulation of FLS2 signaling transduction. The effects of root tissue type and cell identity on FLS2 signaling demonstrate that responses to flg22 are compartmented and modular (Zhou et al., 2020, Emonet et al., 2020, Rich-Griffen et al., 2020). Here, we extend this response complexity by reporting that flg22 variation can differentially activate defense pathway outputs. We find that deviant variants and antagonistic priming variants have negatively charged amino-acids in the N-terminus of the peptide, suggesting that the charge of this region controls signal attenuation. Negatively charged mutations in the C-terminus of the Pa22 peptide affect FLS2-BAK1 complex formation (Parys et al., Chapter 2). Therefore, we speculate that FLS2-BAK1 binding dynamics play a role in signal specificity and modulation. Taken together, regulating expression of FLS2 in specific cell types and sequestering the receptor from chronic stimulation by a subset of commensal-derived receptor antagonists might represent a passive mechanism to ensure appropriate communication with commensal microbiota. How FLS2 is able to mediate tissue specific responses while selectively allowing subsets of flg22 variants to modulate signal transduction is an exciting open question.

FLS2 and signal transduction as a tabulation machine. The integration of flg22 signaling outputs in a dominant background of evasive peptides is fascinating. We propose that defense attenuation via receptor antagonism and signal modulation is layered on top of signal specificity. In this scenario, FLS2 integrates the ratios of immunogenic pathogen-derived danger signals (canonical peptides) and safety signals (evading and modulating peptides) by operating as a molecular tabulator. We envision that during colonization, bacteria producing danger signals that cannot be masked by suppression mechanisms are accommodated by bacteria producing safety signals. In the steady state balanced microbiota, the detection of pathogen over-proliferation is monitored by increases in danger signals and/or decreases in safety signals across a specific threshold. The identification of signal imbalance could be triggered via the increased colonization of a pathogen producing danger signals and/or alternatively through the decrease in absolute abundance of bacteria that produce safety signals. To add to this complexity, plants can alter the threshold needed to produce defense by tuning FLS2 expression in different cell types and developmental states. In the context of damage, which increases FLS2 expression, we hypothesize that plants increase that amount of safety signals needed for colonization. This would allow plants the ability to differentiate between friend and foe.

Here, we show that commensal communities contain substantial flg22 diversity leading to pervasive evasion, specific defense activation, and differential immune system signal modulation. Plants are colonized by bacterial communities encoding a wide range of proteinaceous MAMPs that are presumably detected by distinct immune receptors. Thus, our study of the natural responses of FLS2 to flg22 diversity provides a conceptual framework to better understand how the diversity in other MAMP and receptor pairs impacts the way

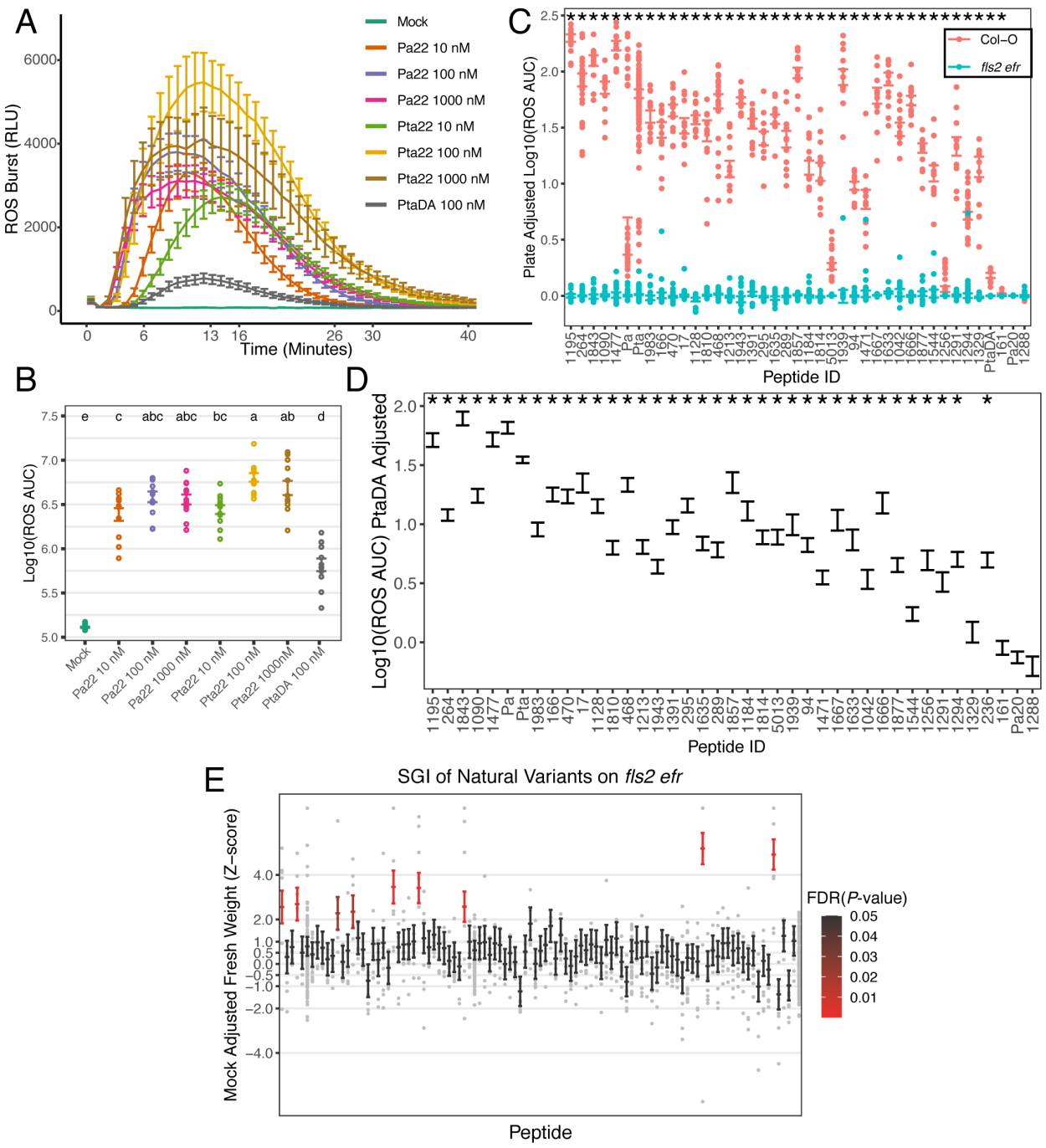
commensal communities are built and maintained to allow for colonization while simultaneously surveilling for pathogens.

Section 3.4: Supplemental Figures



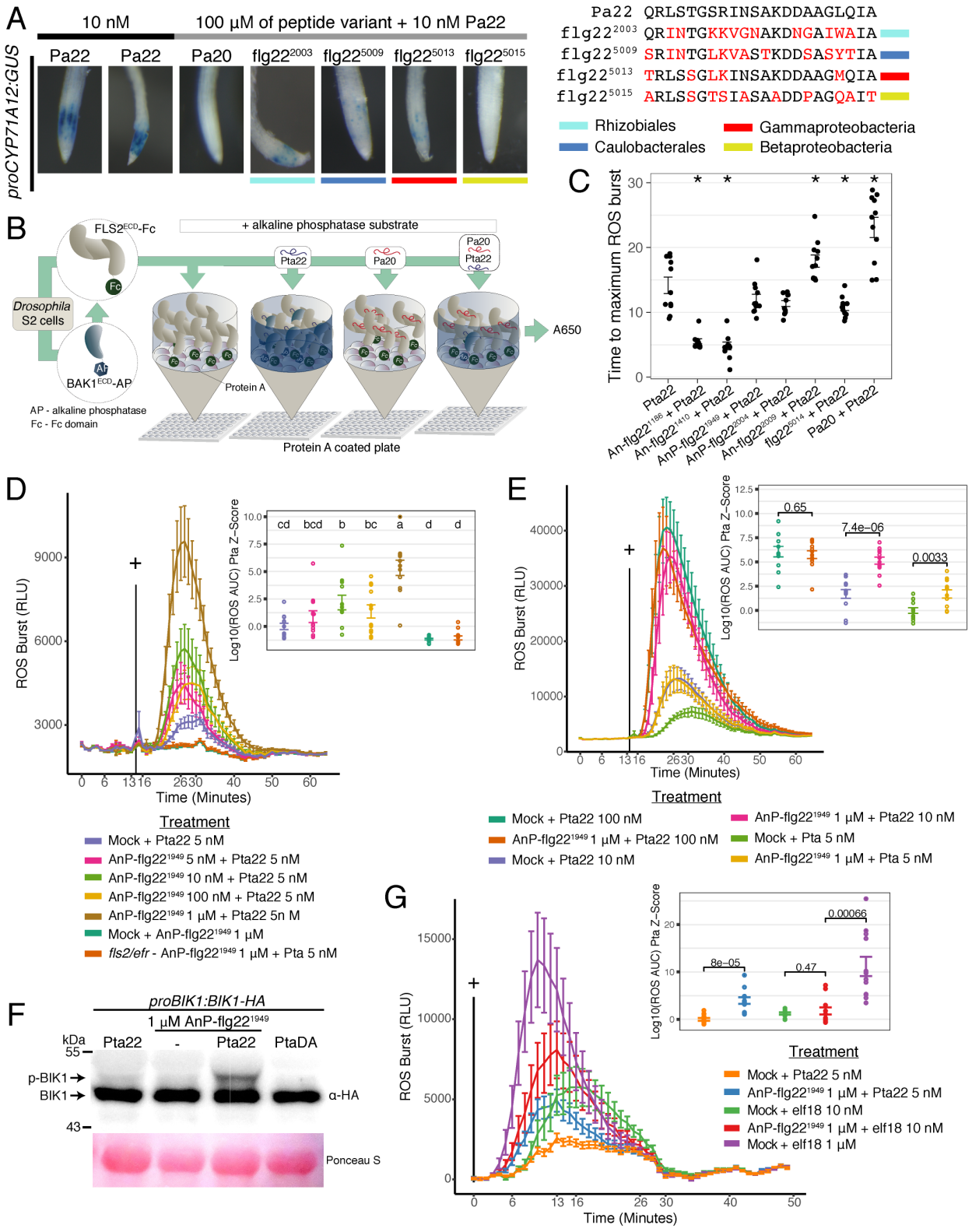
Supplemental Figure 3.8 for Figure 3.1. Multiple copies of *fliC* are found within operons. A) Shown is a zoom-in of the FliC phylogeny from Figure 1A that separates the three clades. There is support for at least three FliC clades, although whether the FliC within clade 2 should be treated as a single clade or as two remain incompletely resolved. For simplicity, we treat it as a

single group given the taxonomic relationships and small number of FliC genes in one of the groups. Blue numbers indicate the branch length and the red numbers indicate the node support based on a Shimodaira-Hasegawa test. The black numbers indicate the number of FliCs, the inside coloring indicates the clade, and the outer coloring indicates the taxonomy of the FliC proteins that emanate from that branch. B) Many *fliC* genes are found in the same operon. Genomic tree shown is the same as Figure 3.1B. Metadata displayed from inner to outer circles represent: taxonomy of bacterial genome, clade of FliC proteins within the genome, the number of *fliC* genes found in a genome, and the number of *fliC* operons in a genome. Operons are defined as groups of *fliC* genes within 10 kB of each other.



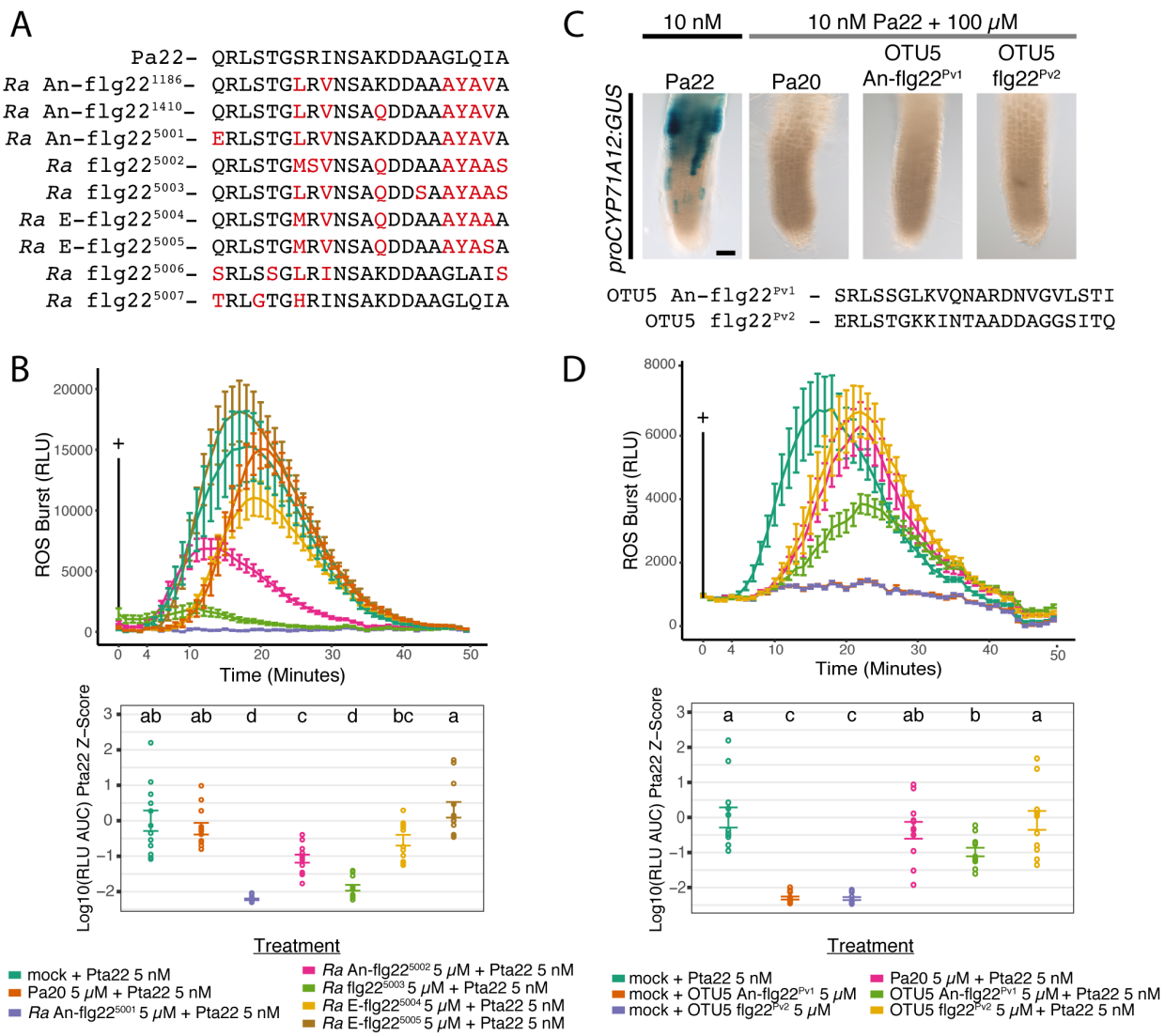
Supplemental Figure 3.9 for Figure 3.2. All ROS-burst and SGI activity produced by flg22 peptide variants are confirmed at a concentration of 10 nM for ROS and are FLS2 dependent. A-B) Pta22 and PtaDA provide a dynamic range to define ROS-burst kinetics of flg22 peptides. Flg22-induced ROS-burst in wild type Col-0 plants driven by the peptides indicated, n=12. Colors are the same in A and B. Distilled water was used as mock control. A) ROS-burst kinetics curves for flg22 peptide variants added at the concentrations shown. Each line represents the mean ROS-burst over time and the error bars indicate the standard error (SE). RLU stands for relative luminescence units. B) Integration of area under the curve (AUC) measurements from the independent leaf disks displayed in A. Bars represent the mean ± the SE. Significantly different

AUC measurements are indicated by letters and were identified using an ANOVA followed by a Tukey's test ($\alpha = 0.05$). C) The ROS-burst kinetics produced by flg22 peptide variants at 100 nM is FLS2-dependent. ROS-burst kinetics produced by Col-0 and *fls2 efr* leaf disks when exposed to 100 nM of each flg22 peptide variant. Each bar represents the median $\pm$ the mean SE for AUC measurements from at least two experimental replicates each with 12 independent leaf disks. Each peptide had independent t-tests performed between Col-0 and *fls2 efr* AUC measurements (* = $P < 0.05$). D) Flg22 peptide variant activity at 100 nM is confirmed at 10 nM. Each flg22 peptide variant was tested on wild type Col-0 leaf disks at 10 nM. Each barplot represents the estimated Log10(AUC) $\pm$ SE difference in ROS-burst to PtaDA from a linear mixed model using at least 24 leaf disks from 2 experimental replicates controlling for plate effect. T-statistics were estimated with Satterwhite approximation and used to calculate *P*-values for the comparison of 10 nM of each peptide to PtaDA. *P*-values were then FDR corrected (* = $P < 0.05$). E) All SGI activity observed for SGI active flg22 peptide variants are FLS2-dependent. Flg22-induced seedling growth inhibition in *fls2 efr* plants. Measurement normalization and statistical analysis are the same as described in Figure 3.2A. The FDR corrected *P*-value is colored on the error bars which represent the estimated effect $\pm$ SE.



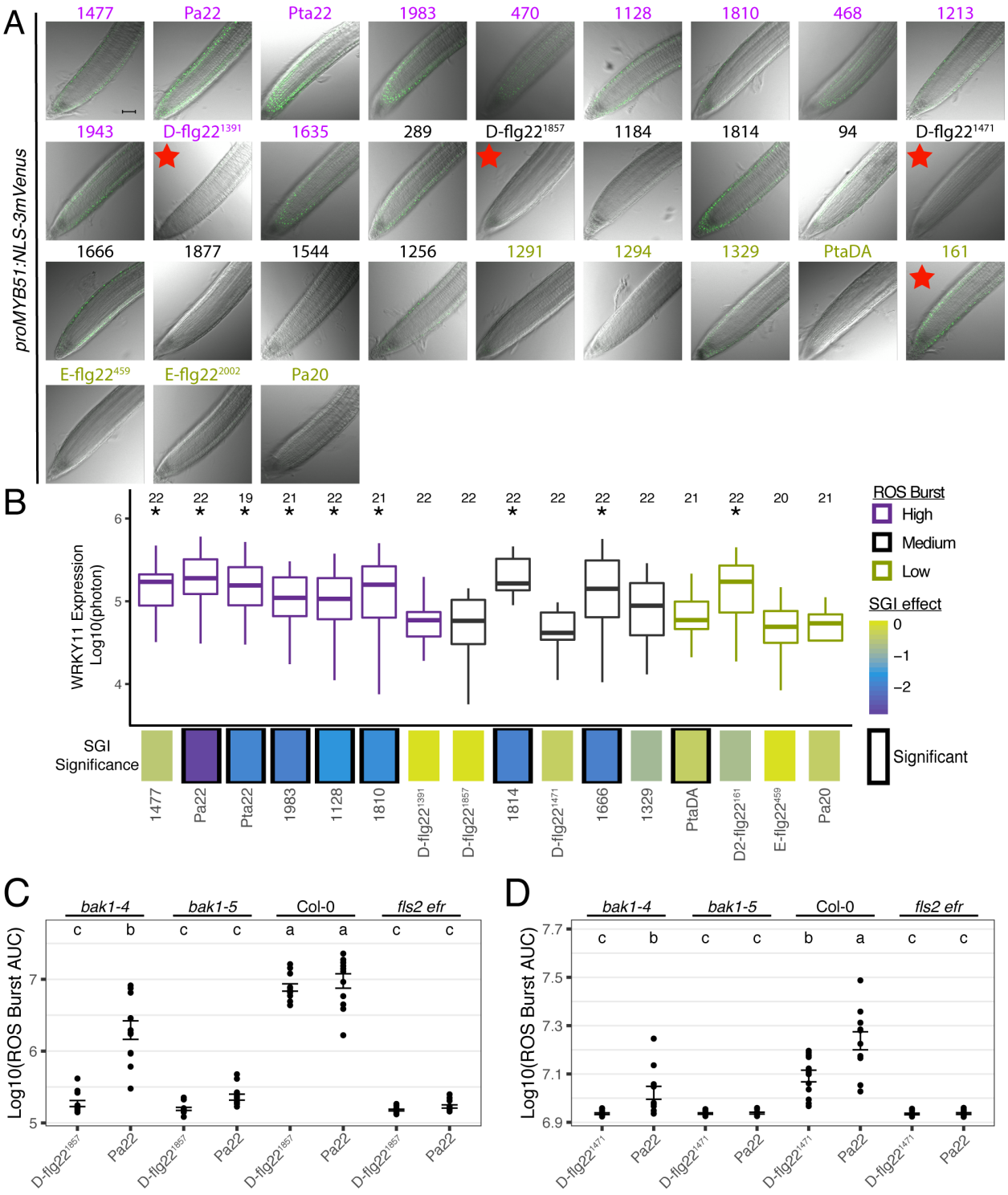
Supplemental Figure 3.10 for Figure 3.3. A novel mechanism of antagonism. A) We identified ten consistent antagonistic peptides within the inactive flg22 variants. 10 nM of Pa22 and 100 µM of each variant flg22 peptide were added to five to eight 7-days-old *proCYP71A12:GUS* seedlings

for five hours before GUS staining. These are the remaining four peptides not shown in Figure 3.4A. Each peptide shown suppressed GUS expression in all roots from two independent experiments. Flg22 variant amino acid sequences are displayed to the right and all amino-acids that differ from Pa22 are colored red. Colors below root pictures and to the right of sequences indicate the taxonomic group that variant is derived from. B) Schematic representation of the ELISA-based FLS2-BAK1 ectodomain interaction assay used to test the ability of antagonists (Pa20) to reduce Pta22 triggered heterocomplex formation. BAK1-conjugated alkaline phosphatase (AP) activity is measured at 650 nm (A650) after application of alkaline phosphatase substrate. This measurement corresponds to the amount of heterocomplex formation. C) Pa20 and flg22^{2009} significantly delays ROS burst induction by Pta22. Each point represents the time where maximum ROS burst was observed after the addition of Pta22, n=11. This is the same experiment displayed in Figures 3.3J-K. Statistical significance was determined using a one-way ANOVA followed by Dunnett's multiple comparison test (* = $P < 0.05$). Bars represent the mean ± the standard error. D-E) Flg22-induced ROS-burst in Col-0 and *fls2 efr* plants. ROS-burst kinetics curves for the Pta22 peptide added 15 min after exposure to AnP-flg22^{1949}. Distilled water was used as a mock control, AnP stands for priming antagonist (see text for description), and RLU stands for relative luminescence units. Each line represents the mean and the error bars represent the standard error of the measurements at each timepoint. Boxplots on the top right are the integration of the area under the curve (AUC). Each boxplot represents the mean ± standard error, n = 12. The experiment was repeated three times with similar results. C) The priming effect of the AnP-flg22^{1949} variant is significant with as little as 10 nM of AnP-flg22^{1949}. Significantly different AUC measurements are denoted by letters and were determined with an ANOVA followed by a Tukey's test ($\alpha = 0.05$). D) The priming effect of AnP-flg22^{1949} cannot exceed the maximum ROS-burst induced by Pta22. *P*-values were calculated with a t-test. F) AnP-flg22^{1949} primes BIK1 phosphorylation. Western blot analyses of BIK1 phosphorylation in *proBIK1:BIK1-HA* seedlings exposed to water or 1 μM of AnP-flg22^{1949} 15 minutes before the addition of 10 nM of Pta22. Plants were vacuum infiltrated for 10 minutes and then left to incubate for an additional 30 minutes before plants were collected for the analysis of BIK1 phosphorylation. Anti-HA antibodies were used to analyze the lysates. The experiment was repeated two times with similar results. Note that 10 nM of Pta22 is not a positive control, but the baseline for BIK1 priming by Pta22 + AnP-flg22^{1949}. G) The priming effect of AnP-flg22^{1949} is flg22 specific. Shown are ROS-burst kinetics curves for Pta22 and elf18 after exposure to 1μM of priming antagonist peptide or distilled water (mock). Each line represents the mean and the bars represent the standard error of the measurements at each timepoint, n = 12. Boxplots on the top right are the integration of the AUC of the ROS-burst curves shown. Each boxplot represents the mean ± standard error. *P*-values were calculated for the comparisons shown using a t-test. The experiment was repeated three times with similar results.



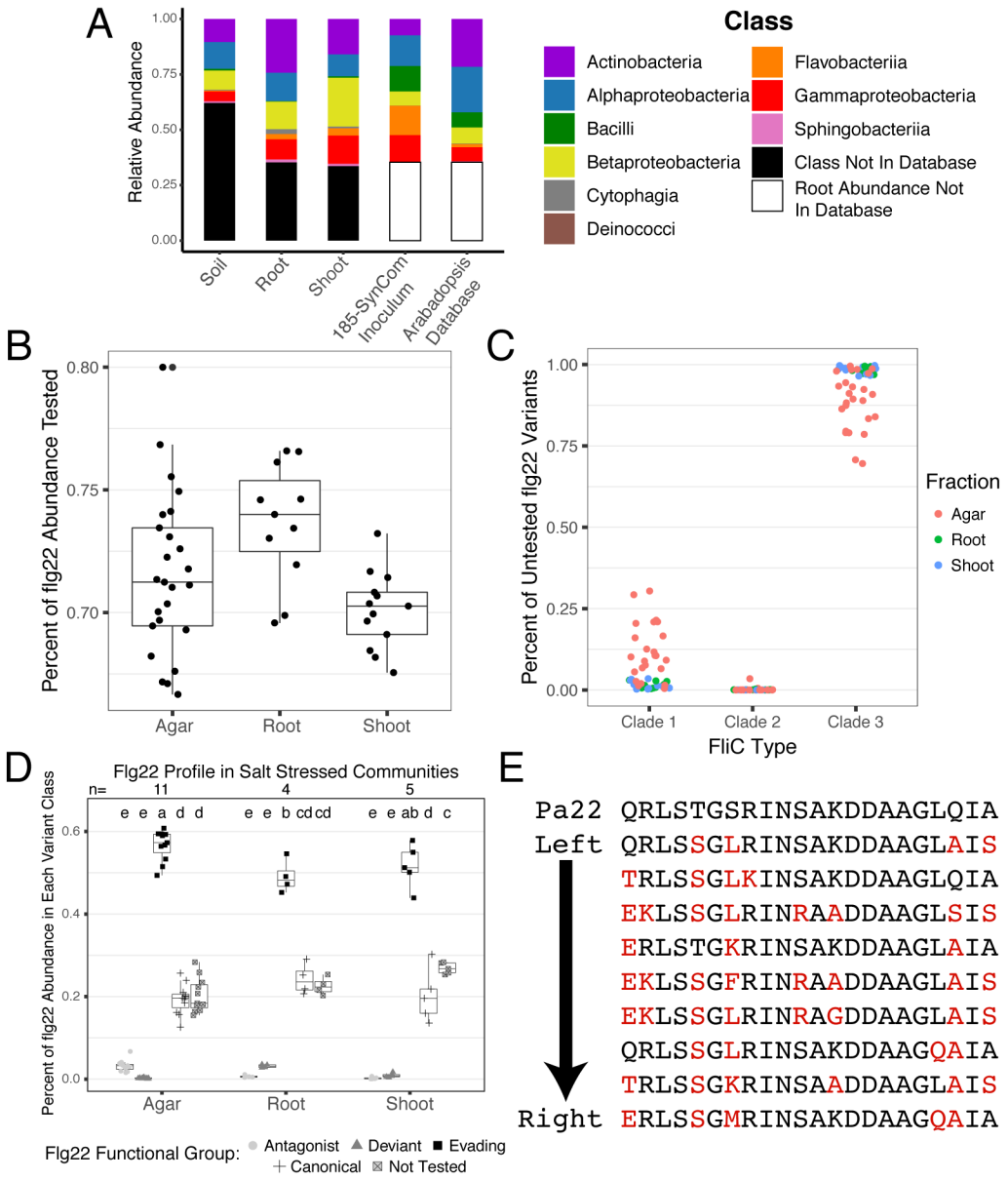
Supplemental Figure 3.11 for Figure 3.4. Prevalent and abundant commensal microbes encode antagonistic flg22 variants. A) *Ralstonia* genomes from NCBI contain nine unique flg22 variants. Positions that differ from Pa22 are marked in red. Note that *Ra* An-flg22^{1186} and *Ra* An-flg22^{1410} were identified in the Arabidopsis commensal database too and are shown in Figure 3.3A. "An-" stands for antagonist and "E-" stands for evader (see text for description). B) Flg22 variants from *Ralstonia* can antagonize ROS induced by 5 nM of Pta22. ROS-burst kinetics curves for Pta22-mediated ROS when exposed with 5 μM of flg22 peptides indicated. Distilled water was used as a mock control, and RLU stands for relative luminescence units. Each line represents the mean and the error bars represent the standard error of the measurements at each timepoint. The boxplot on the top right shows the integration of the area under the curve (AUC) for the ROS-burst kinetics curves. Each boxplot represents the mean ± standard error, n = 12. Significantly different AUC measurements are denoted by letters and were determined with an ANOVA followed by a Tukey's test ($\alpha = 0.05$). C) Both flg22 variants found in *Pv* OTU5 antagonize Pa22-mediated *proCYP71A12:GUS* expression. 10 nM of Pa22 and 100 μM of each variant flg22 peptide were added to four 7-days-old *proCYP71A12:GUS* seedlings for five hours before GUS staining. Scale bar represents 50 μm. Each peptide shown suppressed GUS expression in all roots from two

independent experiments. The sequences of the two OTU5 variants are shown below. D) The OTU5 An-flg22^{Pv1} variant antagonizes ROS induced by 5 nM of Pta22. ROS-burst kinetics curves for Pta22 when exposed with 5 μM of the indicated flg22 peptides. Distilled water was used as a mock control, and RLU stands for relative luminescence units. Each line represents the mean and the error bars represent the standard error of the measurements at each timepoint. The boxplot on the top right shows the integration of the area under the curve (AUC) for the ROS-burst curves. Each error bar represents the mean ± standard error, n = 12. Significantly different AUC measurements are denoted by letters and were determined with an ANOVA followed by a Tukey's test ($\alpha = 0.05$).



Supplemental Figure 3.12 for Figure 3.5. Naturally derived flg22 variants drive separable MTI responses. A) Flg22 peptides produce *MYB51* defense marker gene expression changes in roots. Shown are representative fluorescent, merged brightfield confocal microscope images of pro*MYB51:NLS-3mVenus* roots when exposed to 10 nM of each peptide variant for 24 hours. The red stars in the pictures indicate the peptides highlighted in Figure 3.5A. B) Flg22 peptides induce *WRKY11* defense marker gene expression changes in epidermal cells of cotyledons. Fluorescence

signal quantifications for microscopic observations of *proWRKY11:NLS-3mVenus* expression after being exposed to 10 nM of each flg22 peptide for 24 hours. Peptides are ordered and color-coded by ROS burst activity derived from Figure 3.2A. The number above each boxplot represents the number of roots imaged for each variant. Significant *proWRKY11:NLS-3mVenus* expression was determined by comparing each variant to the gene expression observed for Pa20 using an ANOVA and Dunnet's test. * = $P < 0.05$. C-D) Flg22-induced ROS-burst in Col-0, *fls2 efr*, *bak1*-5, and *bak1*-4 plants measured for D-flg22^{1857} (C) and D-flg22^{1471} (D) is BAK1-dependent. Shown are the integration of the area under the curve (AUC) for the ROS-burst curves induced by D-flg22^{1857} (C), D-flg22^{1471} (D), and Pa22 at 100 nM. Each error bar represents the mean $\pm$ standard error, n = 12. Significantly different AUC measurements are denoted by letters and were determined with an ANOVA followed by a Tukey's test ($\alpha = 0.05$). The experiment was repeated three times with similar results.



Supplemental Figure 3.13 for Figure 3.7. Most flg22 variants found in Clades 1 and 2 in the commensal communities assembled from a 185-member SynCom can be annotated. A) The 185-SynCom and Arabidopsis database roughly recapitulates natural communities. Soil, root, shoot, and 185-SynCom inoculum 16S abundance data as reported by Finkel et al displayed at the class level with the taxonomic composition of the Arabidopsis database (Finkel et al., 2019). All 16S abundance from wild soil, root, and shoot data assigned to a class not found in the Arabidopsis database are represented by black. To improve community comparison, the 185-SynCom inoculum and Arabidopsis taxonomic composition data are scaled to the total abundance of root bacteria not assigned to a class in the Arabidopsis database (35.35%, depicted in white). B) For all the samples analyzed, we functionally annotated about two-thirds of the flg22 abundance in the 185-member SynCom communities on Arabidopsis seedlings grown in agar (Finkel et al., 2020). The relative abundance of each microbe was assigned to the flg22 variant it encoded. Microbes with multiple flg22 variants contributed their relative abundance to the count of each unique flg22 variant in the genome. Percentages were calculated based on the total relative abundance of all flg22 variants in

the community. C) Almost all of the functionally untested flg22 diversity in the 185-member communities come from Clade 3 FliCs. The percentage of untested flg22 variants from each FliC clade in each sample from the experiment are displayed. D) Immunogenic canonical flg22 variants are not depleted and evading variants are not enriched when communities are formed under salt stress. A 185-member SynCom was inoculated onto Arabidopsis seedling with 200 mM NaCl (Finkel et al., 2020). Microbial relative abundance values were assigned to all their flg22 variants, and percentages were calculated using the total flg22 abundance found in the community using the assigned relative abundance values. E) The abundant and prevalent active flg22 amino-acid sequences in the 185-member SynCom presented in Figure 3.7E.

Section 3.5: Methods

Plants. *Arabidopsis thaliana* accession Columbia (Col-0) was used as wild type (WT) in all experiments performed. The mutant lines *fls2 efr* (SAIL_691_C4 / Salk_044334), *bak1-4,* and *bak1-5* were described elsewhere (Nekrasov et al., 2009, Chinchilla et al., 2007, Schwessinger et al., 2011). For all experiments testing the necessity of FLS2, *fls2 efr* mutant plants were used. The EFR and FLS2 receptors are genetically independent, and their respective immunogens are significantly different from one another. It is extremely unlikely that any flg22 variant would signal via EFR, and data shows this to be so (Nekrasov et al., 2009, Zipfel et al., 2006). The transgenic lines *proFLS2:FLS2-GFP* (Chinchilla et al., 2007), *proMYB51:NLS-3mVenus* (Poncini et al., 2017)*, proWRKY11:NLS-3mVenus* (Poncini et al., 2017), *proCYP71A12:GUS* (Millet et al., 2010), *proBIK1:BIK1-HA* (Lin et al., 2014), and *proUBQ10:FLS2-GFP* (Wyrsch et al., 2015) were described previously. Plants were grown on soil in short-day light conditions (12 h light/12 h dark) or vertically in Petri dishes containing 1/2 Murashige and Skoog (MS) medium, 0.8% plant agar, and 1% sucrose in long-day light conditions (16 h light/8 h dark) unless otherwise specified.

Bacteria. The bacterial strains that were not generated for this study include: *Pseudomonas syringae* pv. *tomato* DC3000Δ*avrPtoA/B* (He et al., 2006) and *Ralstonia sp.* UNC404CL21Col (NCBI:txid1380362, henceforth CL21, Levy et al., 2018). The CL21Δ*fliC*,

CL21Δ*fliC::Ra* flg22^{5003}, and CL21Δ*fliC::Ra* flg22^{5005} strains generated in this study were created from the *Ralstonia_sp._UNC404CL21* strain. All bacteria were stored in 40% glycerol stocks at -80°C.

Genomic information. We downloaded all 3837 genomes of plant-associated bacteria from a previously built database (Levy et al., 2018). We used the metadata to identify genomes isolated from Arabidopsis, resulting in 627 genomes (Table S1). Of these, 390 isolates were collected from the root rhizoplane and root endophytic compartments, while the rest are derived from rhizosphere and leaf communities (Table S1). We downloaded 156 *Ralstonia* genomes from NCBI on 12/20/2018 (database on Github at https://github.com/ncolaian/NatVar_proj) and extracted 152 FliC genes that we used for subsequent analyses (Table S4).

FliC identification and phylogenetic tree creation. 387 high quality FliC proteins were selected by first performing psiblast on the 3837 genome Levy et al. database (Altschul et al., 1997, Levy et al., 2018) with the FliC protein from *Pseudomonas aeruginosa*. These results were then filtered based on PFAM and KO annotations. Each protein used is annotated with at least one of the following: flagellin PFAM annotations of pfam00669 and pfam00700, and/or the flagellin KO annotation of K02406. This resulted in 387 FliC proteins (Supplemental Data 1) which we then aligned with MUSCLE (Edgar, 2004). We used the resulting multiple sequence alignment to create a hidden Markov model (HMM) profile with hmmbuild and then scanned each genome in our database using hmmscan from the HMMER3 software package (Johnson et al., 2010). We only considered genes with an e-value equal to or greater than 10^{-50} as encoding FliC proteins in all subsequent analyses. This cutoff produced high specificity for FliC proteins; we found no other genes in our dataset at this threshold.

The resulting proteins were aligned using MUSCLE (Edgar, 2004). The alignment was trimmed to include columns containing amino acids from 90% of the sequences with trimAI (Capella-Gutiérrez et al., 2009), and the protein phylogeny was created using FastTree (Price et al., 2010). Branch length is estimated by FastTree and is interpreted as the substation rate between FliC proteins, and is normalized to the substitution rate of the FliC proteins to randomly generated proteins (Price et al., 2010). We observed, at a minimum, three distinct clades. It is possible that there could be four based on the node support calculated by a Shimodaira-Hasegawa test from FastTree (Price et al., 2010). For simplicity, we collapsed the Bacilli and Actinobacteria FliC clades together since one potential clade contained only 20 FliC proteins (Figure 3.8A). Visualization of the phylogenetic tree was performed with the online tool ITOL (Letunic and Bork, 2019). It is important to note that the node support around the clade separations was much stronger than the support closer to the tips of the tree. For visualization purposes, we did not collapse weakly supported nodes, so please be aware that the relationships around the tips of the FliC tree in Figure 3.1A may not be fully resolved.

The median number of *fliC* genes in genomes encoding clade 3 FliCs is two, while the median number for clades 1 and 2 is one. Bacteria with clade 3 FliCs can encode as many as 7 *fliC* genes (Figure 3.1B). This is consistent with observations in *Bdellovibrio bacteriovorus* which encodes 6 *fliC* genes; all FliC proteins produced from these genes are incorporated into the flagellum (Iida et al., 2009). We found that 94% of genomes with more than one *fliC* gene contained FliC proteins from the same clade, indicating that horizontal gene transfer of *fliC* genes is rare (Figure 3.8B).

Identification of flg22 sequences. Flg22 epitope amino acid sequences were identified from the FliC proteins by first splitting the FliC proteins by clade and aligning each clade to the

P. aeruginosa flg22 sequence using MUSCLE (Edgar, 2004). The FliC proteins identified in *Ralstonia* were also aligned to the *P. aeruginosa* flg22. Then, using a custom script, a continuous 22 amino acid stretch of sequence was obtained based on the beginning (N-terminus) of the alignment of each FliC sequence to flg22 in the multiple sequence alignment. In order to ensure proper alignment, position 14 was checked for aspartic acid (D), as this position is the most conserved residue in the flg22 region based on previously known flg22 sequences. Any flg22 sequence not containing aspartic acid (D) at position 14 was flagged for manual curation or discarded.

Position Weight Matrices (PWM). The flg22 sequences were then used to create a position weight matrix (PWM) using a custom script. The PWM matrices were created using a pseudocount that was adjusted based on equal background amino acid frequencies and the number of flg22 sequences being analyzed. These PWMs were then visualized with the online tool Seq2logo (Thomsen and Nielsen, 2012).

***FliC* Operon Analysis.** We defined an operon as units of *fliC* genes that were within 10 kb of each other. This means that an operon could span over 10 kb, but each *fliC* would be no more than 10 kb from another. They were identified using a custom script.

Peptide Synthesis. All flg22 peptides listed in Table S2 were synthesized at >95% purity by an in-house protein chemistry facility (Gregor Mendel Institute, GMI), Shanghai Apeptide Co., and/or GenScript. All experiments were performed using peptide synthesized from GMI or Shanghai Apeptide Co, except for the leaf infection assay performed using peptides synthesized from GenScript. All synthesized peptides were dissolved in 200 μL of pure water and their concentration was quantified using the Direct Detect Spectrometer (Millipore Sigma). These stock solutions were used to prepare 100 μM aliquots of the peptides and stored at -20 °C. The

purity of all peptides used were validated with MALDI-TOF MS before and after ROS, SGI, GUS, and gene expression assays. Elf18 peptide (SKEKFERTKPHVNVGTIG, Kunze et al., 2004) was ordered from GenScript.

Reactive Oxygen Species (ROS) Burst. ROS burst assays were performed using a luminol based assay (Lozano-Durán and Belkhadir, 2017). 5mm leaf disks were taken from the leaves of healthy 6-week-old Arabidopsis plants and placed adaxial side up in a 96-well plate (Greiner bio-one, ref 655075) with 100µL of distilled water overnight. The distilled water was replaced with a solution of 200µM luminol (SIGMA Cat. No. A8511) and 10µg/mL of Horseradish Peroxidase (Fisher Cat. No. 31490) with or without a flg22 peptide. The plate was immediately placed into the luminometer (SpectraMax L, Molecular Devices) where luminescence was recorded each minute for 50 minutes at the target wavelength of 470nm using an integration time of 0.35 s. The peptide concentrations used in the experiments are specified in the figure legends.

The antagonistic ROS burst was performed as described above. The only difference is the timing of flg22 peptide addition to the plants. 50 µL of reaction solution supplemented with potential antagonist at double the final concentration was added to each well and placed in the luminometer for 15 minutes. After, 50 µL of the Pta flg22 peptide solution at double the final concentration was added to the mixture and placed into the luminometer for 50 minutes. The final concentrations of flg22 peptides are specified in each figure legend.

ROS Data Analysis. Area under the ROS kinetics curve (AUC) was calculated by integrating the area under the ROS burst curve from 0-50 minutes with the sintegral function from the Bolstad2 (1.0-28, Bolstad, 2010) package in R.

In order to compare ROS curves across multiple plates, 12 AUC measurements for loss of function flg22 peptide derived from *Pseudomonas syringae* pv. *tabaci* (PtaDA - TRLSSGLKINSAKADAAGLQIA) (Naito et al., 2008) were recorded on every plate. A z-score was calculated for each leaf disk using the AUC mean and standard deviation from the PtaDA replicates from the plate on which it was measured (Fig. 2A-C and Fig. 6B-K).

For the ROS burst experiments comparing the response of each peptide in *fls2 efr* and Col-0 plants, we paired *fls2 efr* and Col-0 plants on the plate for each peptide. We adjusted the Log10(AUC) measurements from each plate by subtracting all AUC measurements by the median *fls2 efr* value on the plate, n=48 (Fig. S2B). 24 AUC measurements from 2 independent experiments from *fls2 efr* and Col-0 plants exposed to each peptide variant were compared using a t-test with the stat_compare_means function from ggpubr (0.2.5).

For the 10nM ROS burst experiments (Fig. S2C), we could not adjust the plate measurements with the PtaDA based Z-score method because it is inactive at 10nM. We still included PtaDA on every plate, and used a linear mixed model with the model formula:
AUC ~ Name + (1|plate)

Where *AUC* is the Log10 area under the curve values, *Name* is peptide ID, and *(1|plate)* models the random plate effect to compare all the peptides to the PtaDA peptide with a two-sided t-test using the lme4 (1.1-24, Bates et al., 2015) and lmerTest (3.1-1, Kuznetsova et al., 2017) packages in R. *P*-values were adjusted with the Benjamini & Hochberg method (FDR) in R (Benjamini and Hochberg, 1995).

Seedling growth inhibition (SGI) assay. Following a defined protocol (Gómez-Gómez et al., 1999), Col-0 or *fls2 efr* Arabidopsis seeds were surface sterilized using the standard bleach sterilization method and then stratified for two days at 4°C. Seeds were then germinated on ½

MS plates containing 0.8% plant agar and 1% sucrose for five days at 22°C. The seedlings were then transferred to a 48-well plate (Griner Bio one, Cat. No. 677180) containing 1 mL of liquid ½ MS medium supplemented with 10 nM of flg22 peptide. After nine days of growth, the seedlings were patted dry and weighed. Each experiment included 8 seedlings per treatment. A mock control and canonical active flg22 sequence from *Pseudomonas aeruginosa* (Pa22-QRLSTGSRINSAKDDAAGLQIA) were assayed on each plate.

SGI normalization and modeling. To normalize the data points across the experiments, a z-score was calculated for each weight by using the mean and standard deviation of the weights observed for the mock controls on each plate. Each flg22 peptide was tested on least 3 different plates for Col-0 experiments (Fig. 2A-C) and only one plate for *fls2 efr* experiments (Fig. S2D). A linear-mixed model was implemented to identify variants that significantly reduced the growth Arabidopsis compared to the mock controls, while controlling for the still significant plate effect. The model is:

$$Z(weight) \sim ID + Exp + (1|plate\_exp)$$

Z(weight) is the z-score calculated for each weight measurement, *ID* is equal to the variant, *Exp* describes the batch the plate was performed in, and *plate_exp* is a unique ID given to each plate using the lme4 (1.1-24, Bates et al., 2015) package in R. A *P*-value was estimated with a two-sided t-test using the lmerTest (3.1-1, Kuznetsova et al., 2017) package in R, and then adjusted with the Benjamini & Hochberg method (FDR) in R (Benjamini and Hochberg, 1995).

Flg22 tree creation. A maximum likelihood tree was created with a WAG substitution matrix using the MEGA7 software package (Kumar et al., 2016) and the flg22 amino acid sequences. The tree was then visualized with the data layered on the tree using the online software ITOL (Letunic and Bork, 2019).

Antagonism of flg22 driven CYP71A12:GUS expression in Arabidopsis seedlings. 5-8 *proCYP71A12:GUS* plants were germinated per well in a 48-well plate (Griner Bio one, Cat. No. 677180) in liquid MS (Murashige and Skoog basal medium with vitamins containing 0.5 g/L MES hydrate and 0.5% sucrose at pH 5.7; Millet et al., 2010). After 7 days, a solution of liquid MS and antagonist peptide at final concentration of 100 μM was added and pre-incubated with seedlings for 15 minutes. Next, the media was supplemented with 10 nM Pa22 and incubated for 5 hours at 21°C. The media was then removed, and each well was washed with 50 mM sodium phosphate (pH 7). The GUS substrate solution was added (50 mM Na3PO4, 10 mM EDTA, 0.5 mM K4[Fe(CN)6], 0.5 mM K3[Fe(CN)6], 0.5 mM X-Gluc, 0.01 % Silwet L-77) and incubated overnight in the dark at 37°C. Seedlings were then fixed in 3:1 EtOH:acetic acid at 4°C for 5 hours and stored in 95% EtOH. Root pictures were taken using a Leica M205FA stereoscope coupled to a Leica DFC310FX camera.

Protein extraction and Co-immunoprecipitation in Arabidopsis. *proFLS2:FLS2-GFP Arabidopsis thaliana* plants were grown for two weeks on ½ MS agar plates with 1% sucrose in long-day photoperiod. About 0.5 g of plant material was transferred into 6-well plates (Griner Bio one, Cat. No. 657185). 2mL of water supplemented with each flg22 variant was added to the wells and the plants were vacuum infiltrated for 15 minutes. Plants were then dried and ground to a fine powder in liquid nitrogen. The ground plant tissue was then incubated with ~ 2 mL of protein extraction buffer (50 mM Tris pH 7.5, 100 mM NaCl, 10% glycerol, 5 mM EDTA, 1 mM Na_2MoO_4, 20 mM NaF, with fresh 1 mM DTT, and 1 protease inhibitor cocktail tablet/50 mL) and 1% IGEPAL for 30 minutes at 4°C with rotation. The samples were then centrifuged for 10 min at 4°C, 1300 rpm, and the supernatant was filtered through Miracloth. 100μL of the filtered supernatant was mixed with 100 μL of Laemmli sample buffer (2x) as input.

The supernatant was diluted 1:1 in extraction buffer without IGEPAL, and 15 µl of equilibrated GFP-Trap A beads (Chromotek, Cat. No. gta-100) were added to each sample. The samples were incubated for 4 hours at 4°C with rotation. The supernatant was discarded and the beads were washed three times in washing buffer (50 mM Tris pH 7.5 and 100 mM NaCl) for 5 minutes at 4°C with rotation (800g). The elution of the protein complexes was performed in 100µL of Laemmli sample buffer (2x), boiled for 10 min at 95°C. This was subsequently analyzed by SDS-PAGE followed by immunoblotting using anti-GFP-HRP (Thermo Fisher Scientific, Cat. No. A10260) and anti-BAK1 (Agrisera, Cat. No. AS12 1858) antibodies. 45 µL of material from the bead extractions was loaded in the western blot analysis for the Co-IP image.

BIK Phosphorylation Assay. Six 2-week old *proBIK1:BIK1-HA* seedlings grown on ½ MS plates were transferred to 6-well plates (Griner Bio one, Cat. No. 657185) containing 5 mL of sterile water. After 30 min, the water was replaced by elicitor solution at the final concentration described in the figures and legends. For assays with the priming antagonist, a solution containing the priming antagonist was added 15 minutes before the addition of the elicitor. The seedlings were then vacuum infiltrated for 10 minutes and left to incubate for 30 minutes. Immediately afterwards, the seedlings were transferred to 2 mL tubes containing glass beads and frozen with liquid N_2. Tissues were homogenized with TissueLyser II (Qiagen) and 300 µL of protein extraction buffer (50 mM Tris-Cl pH 7.5, 100 mM NaCl, 10 % glycerol, 5 mM EDTA, 1 mM Na_2MoO_4, 20 mM NaF, 1 mM DTT, Complete Mini EDTA-free protease inhibitor cocktail (Roche) was added. After centrifugation (20 min, 20,000 *g*, 4ºC), the supernatant mixed with Laemmli sample buffer and denatured for 5 min at 95°C. The proteins were resolved in 10% SDS-PAGE gels and transferred to nitrocellulose membrane (GE healthcare). HA-tagged

BIK1 proteins were immunoblotted with primary anti-HA (Roche, 1:1000) and secondary anti-rat IgG-HRP (Abcam, 1:5000).

Western Blot Imaging. Western blots were imaged two separate ways. For figures 3.10F and 5B-C, proteins were illuminated using Amersham ECL Prime Western Blotting Detection Reagent (Cytiva, RPN2236). Images were taken by placing the membrane in a dark box and taking pictures with a digital camera using KwikQuant Imager (Kindle Bio). For figure 3.3B, proteins were illuminated using SuperSignal™ West Pico PLUS Chemiluminescent Substrate (Thermo Scientific, REF 34579). Images were taken using KODAK X-OMAT 5000 RA Processor and CL-XPosure™ Film (12,5x17,5 cm) (Fisher Scientific).

***In vitro* FLS2-BAK1 ectodomain interaction studies.** This assay was performed as described in (Parys et al., submitted, see figure 3.10B). Briefly, Extracellular Domains (ECDs) of FLS2 cloned in pECIA2 (bait protein, plate bound) and BAK1 cloned in pECIA14 vector (prey protein, interacts with substrate) were transiently expressed in Drosophila melanogaster Schneider 2 (S2) cells. The protein expression was then confirmed by immunoblotting using anti-V5 (Invitrogen, Cat. No. R961-25) and anti-Flag (Sigma Aldrich, Cat. No. A8592) antibodies. Following, the $FLS2^{ECD}$ was diluted 20 times in 1xPBS containing 0.1 % Tween-20 (PBS-T), and mixed 1:1 with BAK1-ECD diluted 40 times. Accordingly, the Pta22 peptide was added to the protein solution to a final concentration of 10 nM along with the antagonistic flg22 peptide in final concentration ranging from 0.01 – 10 µM. The protein mix was pre-incubated 2 h in 4°C. Following, 100 µl of the protein solution was transferred to a protein-A coated 96 well plate (Thermo Fisher Scientific, ref 15132) and incubated overnight in 4°C. The next day the plate was washed two times with 100 µl of PBS-T and 100 µl of alkaline phosphatase (Sera care, Cat. No. 5120-0059), the substrate was added. Two hours upon addition of the substrate, the

absorbance was measured at 650 nm using a Synergy H4 Multi-Mode plate reader (BioTek). Each plate contained BIR4-BAK1 positive control (Smakowska-Luzan et al., 2018). To control for the prey unspecific binding to protein A coated wells, the bait protein was substituted with a solution of human recombinant IgG1-Fc protein (Invitrogen™ Sino Biological™) at final concentration 625 pg/ul. For each flg22 peptide tested, relative A650 was quantified by dividing the raw A650 after 2h measurement to the signal of mock control (IgG1-Fc).

Microscopic observations of flg22 elicited expression of fluorescent reporter genes. 5-days-old seedlings grown on ½ MS with 1% sucrose agar plates were transferred to liquid medium (½ MS + 1% sucrose) with each flg22 peptide variant at 10 nM. 24hr after treatment, confocal laser scanning microscopy (CLSM) experiments were performed by Zeiss LSM780. Excitation and detection wavelengths were set as 488 nm and 493-598 nm respectively. The adaxial side of leaves for *proWRKY11:NLS-3xmVenus* and meristem zone of roots for *proMYB51:NLS-3xmVenus* were examined by CLSM. Images were processed and analyzed using the Fiji software.

Vector Construction. Vectors were constructed for the deletion of the *Ralstonia sp.* UNC404CL21Col (CL21) *fliC* (pJMC168) and subsequent insertion of various *fliC* alleles containing altered flg22 coding sequences (pJMC174, pJMC176, and pJMC182). All of these vectors were constructed by Gibson Assembly using HiFi DNA Assembly Mastermix (New England Biolabs) to assemble DNA fragments amplified using the Q5 DNA Polymerase including the optional GC enhancer (New England Biolabs). DNA fragments amplified from vector templates were treated with DpnI (New England Biolabs) to digest vector template DNA prior to inclusion in Gibson Assembly reactions, as appropriate. All DNA fragments were cleaned-up as necessary using the QIAquick PCR Purification Kit (Qiagen). All assembled

vectors were transformed into NEB 5-alpha chemically competent *E. coli* (New England Biolabs) and selected on LB media with 30 μg/mL chloramphenicol. Vectors were miniprepped using the ZR Plasmid Miniprep Classic Kit (Zymo Research) and sequence confirmed by Sanger Sequencing (Genewiz). All primer sequences and vectors used in this study are available in Table S3.

Because CL21 is naturally resistant to kanamycin, the antibiotic resistance marker on suicide vector pMo130 (Hamad et al., 2009) was changed from kanamycin resistance to chloramphenicol resistance. Vector pMo130-cmR was assembled using two fragments: the pMo130 vector amplified using primers JMC661-JMC662, and the chloramphenicol resistance gene and its promoter amplified from pLysS using primers JMC663-JMC664. This new vector, pMo130-cmR, was used as the base vector for subsequent vectors. For the knockout vector pJMC168, regions flanking *fliC* were amplified from CL21 genomic DNA using primers JMC635-JMC636 for the upstream flanking region and primers JMC637-JMC638 for the downstream flanking region. These were assembled with the pMo130-cmR backbone amplified with primers JMC203-JMC204. To construct a base vector for the *fliC* variant knock-in vectors (pJMC174), primers JMC635-JMC638 were used to amplify the region containing the wild type *fliC* and both flanking regions. This was assembled with the pMo130-cmR backbone amplified with primers JMC203-JMC204 resulting in pJMC174. The *fliC* variant knock-in vectors (pJMC176 and pJMC182) were constructed by assembling a 450bp DNA synthesis product (Genewiz), starting 3bp before the *fliC* start codon and containing an altered flg22 coding sequence in *fliC* (*Ra* E-flg22^{5005} in pJMC176 and *Ra* flg22^{5003} in pJMC182), with the vector backbone amplified from pJMC174 using primers JMC719-JMC720. The DNA synthesis product sequences for these vectors can be found in Supplementary Table 3. To prepare for

biparental mating with CL21, all vectors were transformed into biparental mating *E. coli* strain WM3064 and selected on LB agar plates containing 30 µg/mL chloramphenicol and 0.3 mM diaminopimelic acid (DAP) at 37°C.

CL21*ΔfliC* strain construction. The unmarked deletion of *fliC* (Gene ID 2558853875) in strain *Ralstonia sp.* UNC404CL21Col (CL21) was constructed using two-step allelic exchange based on a genetic system developed for *Burkholderia* spp. (Hamad et al., 2009). Biparental mating between *E. coli* strain WM3064 containing knockout vector pJMC168 and CL21 was performed by growing each strain overnight: *E. coli* containing pJMC168 in LB medium with 30 µg/mL chloramphenicol and 0.3 mM diaminopimelic acid (DAP) at 37°C and CL21 in 2xYT medium containing 100µg/mL ampicillin at 28°C. The strains were washed three times with 2xYT medium lacking antibiotics, resuspended in 1/10 the volume and mixed in equal proportion donor:recipient, and plated on LB agar containing 0.3 mM DAP and grown overnight at 28°C. Exconjugates were selected on LB agar plates containing 150 µg/mL chloramphenicol, 100µg/mL ampicillin, and lacking DAP at 28°C. First crossover strains were confirmed for the insertion of knockout vector pJMC168 at the correct genomic location in CL21 exconjugant strains using primers outside the regions of homologous recombination (JMC655, JMC656) and primers in the pMo130-cmR vector backbone (JMC321, JMC634). Confirmed first crossovers were resolved by passaging one time on 2xYT medium containing 100µg/mL ampicillin and 1mM IPTG. This was followed by one passage on media containing 10g/L tryptone, 5 g/L yeast extract, 100 g/L sucrose, 100µg/mL ampicillin, and 1mM IPTG, and then plating the strains on the same media containing 1.5% agar. The resulting strains were screened for the deletion of *fliC* using primers JMC655-JMC656. Deletion strains were plate-purified by streaking two additional times on LB agar plates containing 100µg/mL ampicillin. The resulting strain was confirmed by

PCR using JMC655-JMC656 and by confirming the loss of motility using the motility assay described below. The final strain is designated CL21Δ*fliC*.

CL21 *fliC* allele knock-in. The same two-step allelic exchange procedure used to construct the CL21Δ*fliC* strain was used to insert different alleles of *fliC* containing altered flg22 coding sequences into the CL21Δ*fliC* strain background. Biparental matings of *E. coli* WM3064 strains containing vector pJMC176 or pJMC182 were performed with strain CL21Δ*fliC.* First crossovers exconjugants were selected on LB agar plates containing 150 µg/mL chloramphenicol, 100µg/mL ampicillin, and lacking DAP at 28°C. The vector insertion was resolved as described above and the resulting strains were screened by PCR with primers JMC655-JMC656. Strains with a *fliC* insertion were passaged two times on LB agar plates containing 100µg/mL ampicillin before performing PCR confirmation with primers JMC655-JMC656. This PCR product was sequenced to confirm the insertion of the correct *fliC* allele. The resulting strains using the pJMC176 and pJMC182 vectors were named CL21Δ*fliC::Ra* E-flg22^{5005} and CL21Δ*fliC::Ra* flg22^{5003} respectively based on the swapped in flg22 amino acid sequences displayed in Fig. S6A.

Motility Assay. To assess the motility of strains, LB medium plates solidified with 0.3% (w/v) agar were prepared. To prepare an inoculum for the motility assay, strains were grown overnight in 2xYT medium. The motility plates were inoculated by dipping a sterile toothpick into the overnight culture and stabbing it once into the center of the LB 0.3% agar plate. Plates were incubated at 28°C for four days and images taken on a document scanner. Motility distance was quantified as the farthest point of bacterial spread from the inoculation point using ImageJ.

Root Growth Inhibition (RGI) assay. *proUBQ10:FLS2-GFP* seeds were sterilized by vigorously shaking with a 60% bleach solution for 7 minutes followed by washing with sterilized

water four times. Seeds were stratified in the dark at 4°C for 48 hr, and then placed onto sterile Murashige & Skoog (MS) agar plates (M404; Phytotech labs, Lenexa, Kansas) containing 0.5% sucrose. The seedlings were then grown vertically for seven days in a 16-hr light, 21°C / 8-hr dark, 18°C regime.

Bacteria were grown in the dark on Luria Broth (LB) agar supplemented with 100 μg/mL ampicillin for two days at 28°C. Single colonies were selected and grown overnight in 2xYT liquid media at 28°C while being shaken vigorously. The next day, bacteria were centrifuged, then washed three times, and diluted to OD_{600} 0.0002 using 10 mM $MgCl_2$. Bacteria were then spread onto MS plates containing no sucrose with or without 100 nM Pa22. Six 7-day old *proUBQ10:FLS2-GFP* seedlings were transferred onto each plate, and the root length at time of transfer was demarcated. The plants were then grown vertically for seven more days at the conditions above. Plates were scanned with a document scanner, and root elongation measured using a ruler. Significant differences between bacterial treatments were identified using a linear mixed model with lme4 (1.1-24) that controlled for a significant random plate effect and a significant interaction between batch and bacterial treatment (Bates et al., 2015).

Leaf infection assays. We followed a previously described protocol (Chung et al., 2014). The day before bacterial inoculation, water or 100nM solution of the flg22 peptide variant indicated was infiltrated into leaves of 5-week-old Col-0 Arabidopsis plants using a needless syringe. *Pseudomonas syringae* pv. *tomato* DC3000Δ*avrPtoA/B* was plated on King's B medium (KB) supplemented with 50 μg/mL rifampicin and 50 μg/mL kanamycin. The bacteria were then grown overnight in the dark at 28°C. The next day, 1 mL of 10 mM $MgCl_2$ was added to the bacterial plates and carefully scraped to collect bacteria. The optical density (OD) was immediately measured at OD^{600} and a solution of 1 x 10^5 cfu/mL (OD = 0.0002) of

Pseudomonas syringae pv. *tomato* DC3000Δ*avrPtoA/B* and 10 mM $MgCl_2$ was infiltrated into previously pretreated leaves using a needless syringe. Leaves were allowed to dry (~3 hours), and then placed in short day conditions (8h light, 16h dark) at 21°C for 3 days.

After, four 5 mm leaf disks were taken from four separate plants and placed into a sterile 2 mL Eppendorf tube containing three 4mm beads and 400 μL of distilled water. The plant tissue was then lysed for 45 seconds using the FastPrep-24™ (MP Biomedicals). We then added 600 μL of distilled water to each sample, and performed 10x serial dilutions (20 μL in 180 μL of distilled water). 5 μL from each dilution (0-6) were plated on KB plates supplemented with 50 μg/mL rifampicin and 50 μg/mL kanamycin. These plates were placed in the dark at 28°C for ~24-hr and then CFUs were counted under a microscope. We determined statistically different groups using a two-way ANOVA controlling for batch effect and a post-hoc Tukey test as implemented by the HSD.test function in the agricolae (1.3-2) package.

16S community analysis. We downloaded 16S data reported by a previous study that establishes a 185-member SynCom (Finkel et al., 2019). Briefly, soil was collected from the long-term Pi fertilization field ("Field D") trial at the Julius Kühn Experimental Station at Martin Luther University of Halle-Wittenberg (51°29′45.6″N, 11°59′33.3″E). The 16S community composition was determined by forming amplicon sequence variants (ASVs) and assigning taxonomy to these ASVs. The soil, root, and shoot samples are from the collected soil samples (Figure 3.13A). The analysis of the 185-SynCom inoculum sample was performed by mapping 16S reads to each unique 16S sequence variant (Useq) present in the 185-member SynCom (Finkel et al., 2019). For ease of comparison, the inoculum and Arabidopsis database stacked bar figures were normalized to the wild root abundance assigned to classes present in the Arabidopsis database (Figure 3.13A).

FliC and flg22 Analysis of 185-member SynCom. We assessed how variation in FliC and flg22 could affect and be affected by community flg22 composition. We downloaded 16S sequence data generated from a study of bacterial colonization of the plant using an inoculation of a 185-member synthetic community (SynCom) on 7-day old Col-0 seedlings (Finkel et al., 2020). The community was allowed to colonize seedlings for 12 days under a 16-hr dark/8-hr light regime at 21°C light/18°C dark. We reanalyzed data from control plates used in the 1000 μM Pi or 21°C conditions and from plates supplemented with 200 mM of NaCl as the salt stressed condition. For all analyses we used the relative abundance values for each unique sequence variant (Useq) reported by the authors (Finkel et al., 2020). Each Useq is a unique V3-V4 16S variant that is associated with one or more strains in the 185 SynCom.

We first determined the approximate number of each clade of FliC in each community. To estimate the relative abundance of genomes containing each FliC clade (Fig. 7A), we assigned the relative abundance value of each Useq to the FliC clades found within each Useq. If a Useq was associated with multiple FliC clades or had genomes with or without a *fliC* gene, the Useq relative abundance values were assigned to all the FliC clades present and/or to the no FliC group. Thus, a Useq could contribute to the relative abundance of multiple FliC clades. However, even if an Useq contained genomes with multiple FliCs from the same clade, the relative abundance value was counted once for each FliC clade. The relative abundance values for each FliC clade were then tabulated for each sample. Statistically significant differences in relative abundance values for FliC clades within and between plant fractions were identified using a two-way ANOVA, followed by a post-hoc Tukey test using the HSD.test function as implemented by the agricolae (1.3-2) package in R. Our model controlled for a significant interaction between the relative abundance of bacteria and the two different conditions/batches.

We next assessed the functional flg22 repertoire within the communities by assigning each unique flg22 variant within a Useq the relative abundance of that Useq. In principle, this means that a Useq could contribute its relative abundance to as many unique flg22 variants as are harbored within that Useq. To be conservative, the relative abundance values were only counted once for each flg22 variant found within the Useq even if it contained multiple copies of the same flg22 variant. The ratio of un-tested flg22 variants was determined by dividing the total abundance of the flg22 variants not in our screen by the total flg22 abundance we calculated.

$$f_{untested} = \frac{\sum x_{unobserved}}{\sum \hat{x}_{flg22}}$$

Where x and $\hat{x}$ represent relative abundance values for unobserved and all flg22 variants respectively. We calculated the percentage of the untested abundance from each FliC clade by assigning the relative abundance of each flg22 variant to the FliC clade it is found in (Fig. S6B). The functional flg22 repertoire was summarized from each flg22 variant's relative abundance that fell into the five functional classes, Antagonist, Deviant, Evading, Canonical, and Not tested (definitions in this manuscript). Similarly to above, relative abundance values for all flg22 variants in a functional class was divided by the total flg22 relative abundance to get the percentage of the total flg22 relative abundance associated with each functional class.

Flg22 amino-acid sequence logos corresponding to each experimental condition were created by representing each flg22 sequence proportional to its median relative abundance value across all control conditions. The median relative abundance values calculated over all control samples analyzed were multiplied by 1000 and rounded to the nearest integer. This value reflects the number of times this flg22 amino-acid sequence was represented in the complete flg22 functional repertoire list. This list was then used to create PWM matrices and logos.

The analysis of individual flg22 variants present in each sample was performed on a subset of active flg22 variants. The flg22 percent relative abundance values were used to identify active flg22 variants that had percent relative abundance values significantly different from zero using a linear model within the agar, root, and shoot fractions. We plotted the percent of flg22 abundance over all samples for each of the flg22 variants that were significantly different to zero in any of the fractions (Fig. 7E).

Programs used for statistical analysis and data visualization. GraphPad PRISM 8.0 and the R programming environment (version R 3.6.2) were used for data analysis and visualization. Unless specified otherwise, boxplots represent the 1st and 3rd quartiles, and the median is represented in the center. Whiskers indicate the lesser of either the maximum/minimum data points or 1.5*IQR from the 1st and 3rd quartiles.

All statistical tests were performed in R. Tukey tests were performed with the HSD.test function from the agricolae (1.3-2) package. Dunnets tests were performed with the glht() function from the multcomp (1.4-12) package (Hothorn et al., 2008). All linear mixed models were performed using lme4 (1.1-24, Bates et al., 2015). *P*-values from linear mixed models were calculated using Satterthwaite approximation implemented by lmerTest (3.1-, Kuznetsova et al., 2017). When many statistical tests were performed on the same data and many *P*-values were obtained we performed Benjamini and Hochbergs *P*-value correction method (FDR) to control for false discovery rate (Benjamini and Hochberg, 1995). An alpha value of 0.05 was used for all statistical tests to assign significance unless otherwise specified. All statistical analysis methods are described in the figure legends.

REFERENCES

Albrecht, C., Boutrot, F., Segonzac, C., Schwessinger, B., Gimenez-Ibanez, S., Chinchilla, D., Rathjen, J.P., Vries, S.C.D., and Zipfel, C. (2011). Brassinosteroids inhibit pathogen-associated molecular pattern-triggered immune signaling independent of the receptor kinase BAK1. Proceedings of the National Academy of Sciences *109*, 303–308.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research *25*, 3389–3402.

Amin, I.M., Richmond, G.E., Sen, P., Koh, T.H., Piddock, L.J.V., and Chua, K.L. (2013). A Method for generating marker-less gene deletions in multidrug-resistant Acinetobacter baumannii. BMC Microbiology *13,* 158.

Asai, T., Tena, G., Plotnikova, J., Willmann, M.R., Chiu, W.-L., Gómez-Gómez, L., Boller, T., Ausubel, F.M., and Sheen, J. (2002). MAP kinase signalling cascade in Arabidopsis innate immunity. Nature *415*, 977–983.

Bai, Y., Müller, D.B., Srinivas, G., Garrido-Oter, R., Potthoff, E., Rott, M., Dombrowski, N., Münch, P.C., Spaepen, S., Remus-Emsermann, M., et al. (2015). Functional overlap of the Arabidopsis leaf and root microbiota. Nature *528*, 364–369.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software *67*, 1–48.

Bauer, Z., Gómez-Gómez, L., Boller, T., and Felix, G. (2001). Sensitivity of Different Ecotypes and Mutants of Arabidopsis thaliana toward the Bacterial Elicitor Flagellin Correlates with the Presence of Receptor-binding Sites. Journal of Biological Chemistry *276*, 45669–45676.

Belkhadir, Y., Jaillais, Y., Epple, P., Balsemão-Pires, E., Dangl, J.L., and Chory, J. (2011). Brassinosteroids modulate the efficiency of plant immune responses to microbe-associated molecular patterns. Proceedings of the National Academy of Sciences *109*, 297–302.

Belkhadir, Y., Yang, L., Hetzel, J., Dangl, J.L., and Chory, J. (2014). Wiring and Rewiring in Signal Transduction The growth-defense pivot: crisis management in plants mediated by LRR-RK surface receptors. Trends in Biochemical Sciences *39*, 447–456.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological) *57*, 289–300.

Berens, M.L., Wolinska, K.W., Spaepen, S., Ziegler, J., Nobori, T., Nair, A., Krüler, V., Winkelmüller, T.M., Wang, Y., Mine, A., et al. (2019). Balancing trade-offs between biotic and abiotic stress responses through leaf age-dependent variation in stress hormone cross-talk. Proceedings of the National Academy of Sciences *116*, 2364–2373.

Bolstad, W.M. (2010). Understanding computational Bayesian statistics (Hoboken: Wiley).

Boutrot, F., and Zipfel, C. (2017). Function, Discovery, and Exploitation of Plant Pattern Recognition Receptors for Broad-Spectrum Disease Resistance. Annual Review of Phytopathology *55*, 257–286.

Buscaill, P., Chandrasekar, B., Sanguankiattichai, N., Kourelis, J., Kaschani, F., Thomas, E.L., Morimoto, K., Kaiser, M., Preston, G.M., Ichinose, Y., et al. (2019). Glycosidase and glycan polymorphism control hydrolytic release of immunogenic flagellin peptides. Science *364*, 364–374.

Capella-Gutierrez, S., Silla-Martinez, J.M., and Gabaldon, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics *25*, 1972–1973.

Carlström, C.I., Field, C.M., Bortfeld-Miller, M., Müller, B., Sunagawa, S., and Vorholt, J.A. (2019). Synthetic microbiota reveal priority effects and keystone strains in the Arabidopsis phyllosphere. Nature Ecology & Evolution *3*, 1445–1454.

Chinchilla, D., Bauer, Z., Regenass, M., Boller, T., and Felix, G. (2006). The Arabidopsis Receptor Kinase FLS2 Binds flg22 and Determines the Specificity of Flagellin Perception. The Plant Cell *18*, 465–476.

Chinchilla, D., Zipfel, C., Robatzek, S., Kemmerling, B., Nürnberger, T., Jones, J.D.G., Felix, G., and Boller, T. (2007). A flagellin-induced complex of the receptor FLS2 and BAK1 initiates plant defence. Nature *448*, 497–500.

Chung, E.-H., El-Kasmi, F., He, Y., Loehr, A., and Dangl, J.L. (2014). A Plant Phosphoswitch Platform Repeatedly Targeted by Type III Effector Proteins Regulates the Output of Both Tiers of Plant Immune Receptors. Cell Host & Microbe *16*, 484–494.

Chu, H., and Mazmanian, S.K. (2013). Innate immune recognition of the microbiota promotes host-microbial symbiosis. Nature Immunology *14*, 668–675.

Clarke, C.R., Chinchilla, D., Hind, S.R., Taguchi, F., Miki, R., Ichinose, Y., Martin, G.B., Leman, S., Felix, G., and Vinatzer, B.A. (2013). Allelic variation in two distinctPseudomonas syringaeflagellin epitopes modulates the strength of plant immune responses but not bacterial motility. New Phytologist *200*, 847–860.

Edgar, R.C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics *5*, 113.

Emonet, A., Zhou, F., Vacheron, J., Heiman, C.M., Tendon, V.D., Ma, K.-W., Schulze-Lefert, P., Keel, C., and Geldner, N. (2020). Spatially Restricted Immune Responses Allow for Root Meristematic Activity During Bacterial Colonisation, https://doi.org/10.1101/2020.08.03.233817

Felix, G., Duran, J.D., Volko, S., and Boller, T. (1999). Plants have a sensitive perception system for the most conserved domain of bacterial flagellin. The Plant Journal *18*, 265–276.

Finkel, O.M., Salas-González, I., Castrillo, G., Conway, J.M., Law, T.F., Teixeira, P.J.P.L., Wilson, E.D., Fitzpatrick, C.R., Jones, C.D., and Dangl, J.L. (2020). A single bacterial genus maintains root growth in a complex microbiome. Nature. https://doi.org/10.1038/s41586-020-2778-7

Finkel, O.M., Salas-González, I., Castrillo, G., Spaepen, S., Law, T.F., Teixeira, P.J.P.L., Jones, C.D., and Dangl, J.L. (2019). The effects of soil phosphorus content on plant microbiota are driven by the plant phosphate starvation response. PLOS Biology *17*. https://doi.org/10.1371/journal.pbio.3000534

Fliegmann, J., and Felix, G. (2016). Immunity: Flagellin seen from all sides.Nature Plants*2*, 1–2.

Fujii, M., Shibata, S., and Aizawa, S.-I. (2008). Polar, Peritrichous, and Lateral Flagella Belong to Three Distinguishable Flagellar Families. Journal of Molecular Biology *379*, 273–283.

Garrido-Oter, R., Nakano, R.T., Dombrowski, N., Ma, K.-W., AgBiome Team, McHardy, A.C., and Schulze-Lefert, P. (2018). Modular Traits of the Rhizobiales Root Microbiota and Their Evolutionary Relationship with Symbiotic Rhizobia. Cell Host & Microbe *24*, 155–167.

Gómez-Gómez, L., Felix, G., and Boller, T. (1999). A single locus determines sensitivity to bacterial flagellin in Arabidopsis thaliana. The Plant Journal *18*, 277–284.

Hacquard, S., Spaepen, S., Garrido-Oter, R., and Schulze-Lefert, P. (2017). Interplay Between Innate Immunity and the Plant Microbiota. Annual Review Of Phytopathology Interplay *55*, 565–589.

Haiko, J., and Westerlund-Wikström, B. (2013). The Role of the Bacterial Flagellum in Adhesion and Virulence. Biology *2*, 1242–1267.

Hamad, M.A., Zajdowicz, S.L., Holmes, R.K., and Voskuil, M.I. (2009). An allelic exchange system for compliant genetic manipulation of the select agents Burkholderia pseudomallei and Burkholderia mallei. Gene *430*, 123–131.

He, P., Shan, L., Lin, N.-C., Martin, G.B., Kemmerling, B., Nürnberger, T., and Sheen, J. (2006). Specific Bacterial Suppressors of MAMP Signaling Upstream of MAPKKK in Arabidopsis Innate Immunity. Cell *125*, 563–575.

Hothorn, T., Bretz, F., and Westfall, P. (2008). Simultaneous Inference in General Parametric Models. Biometrical Journal *50*, 346–363.

Iida, Y., Hobley, L., Lambert, C., Fenton, A.K., Sockett, R.E., and Aizawa, S.-I. (2009). Roles of Multiple Flagellins in Flagellar Formation and Flagellar Growth Post Bdelloplast Lysis in Bdellovibrio bacteriovorus. Journal of Molecular Biology *394*, 1011–1021.

Johnson, L.S., Eddy, S.R., and Portugaly, E. (2010). Hidden Markov model speed heuristic and iterative HMM search procedure. BMC Bioinformatics *11*.

Kadota, Y., Sklenar, J., Derbyshire, P., Stransfeld, L., Asai, S., Ntoukakis, V., Jones, J.D., Shirasu, K., Menke, F., Jones, A., et al. (2014). Direct Regulation of the NADPH Oxidase RBOHD by the PRR-Associated Kinase BIK1 during Plant Immunity. Molecular Cell *54*, 43–55.

Karasov, T.L., Almario, J., Friedemann, C., Ding, W., Giolai, M., Heavens, D., Kersten, S., Lundberg, D.S., Neumann, M., Regalado, J., et al. (2018). Arabidopsis thaliana and Pseudomonas Pathogens Exhibit Stable Associations over Evolutionary Timescales. Cell Host & Microbe *24*, 168–179.

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Molecular Biology and Evolution *33*, 1870–1874.

Kunze, G., Zipfel, C., Robatzek, S., Niehaus, K., Boller, T., and Felix, G. (2004). The N Terminus of Bacterial Elongation Factor Tu Elicits Innate Immunity in Arabidopsis Plants. The Plant Cell *16*, 3496–3507.

Kuznetsova, A., Brockhoff, P.B., and Christensen, R.H.B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. Journal of Statistical Software *82*, 1–26.

Kühn, M.J., Schmidt, F.K., Farthing, N.E., Rossmann, F.M., Helm, B., Wilson, L.G., Eckhardt, B., and Thormann, K.M. (2018). Spatial arrangement of several flagellins within bacterial flagella improves motility in different environments. Nature Communications *9*, 1–12.

Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Research *47*.

Levy, A., Gonzalez, I.S., Mittelviefhaus, M., Clingenpeel, S., Paredes, S.H., Miao, J., Wang, K., Devescovi, G., Stillman, K., Monteiro, F., et al. (2017). Genomic features of bacterial adaptation to plants. Nature Genetics *50*, 138–150.

Li, X., Lin, H., Zhang, W., Zou, Y., Zhang, J., Tang, X., and Zhou, J.-M. (2005). Flagellin Induces Innate Immunity in Nonhost Interactions That Is Suppressed by Pseudomonas Syringae Effectors. Proceedings of the National Academy of Sciences *102*, 12990–12995.

Lin, W., Li, B., Lu, D., Chen, S., Zhu, N., He, P., and Shan, L. (2014). Tyrosine phosphorylation of protein kinase complex BAK1/BIK1 mediates Arabidopsis innate immunity. Proceedings of the National Academy of Sciences *111*, 3632–3637.

Lozano-Durán, R., and Belkhadir, Y. (2017). A Technical Framework for Studying the Signaling Nexus of Brassinosteroids and Immunity. Methods in Molecular Biology Brassinosteroids 49–61.

Lu, X., Tintor, N., Mentzel, T., Kombrink, E., Boller, T., Robatzek, S., Schulze-Lefert, P., and Saijo, Y. (2009). Uncoupling of sustained MAMP receptor signaling from early outputs in an Arabidopsis endoplasmic reticulum glucosidase II allele. Proceedings of the National Academy of Sciences *29*, 22522–22527.

Meindl, T., Boller, T., and Felix, G. (2000). The Bacterial Elicitor Flagellin Activates Its Receptor in Tomato Cells According to the Address–Message Concept. The Plant Cell *12*, 1783–1794.

Millet, Y.A., Danna, C.H., Clay, N.K., Songnuan, W., Simon, M.D., Werck-Reichhart, D., and Ausubel, F. (2010). Innate Immune Responses Activated in Arabidopsis Roots by Microbe-Associated Molecular Patterns. The Plant Cell *22*, 973–990.

Mueller, K., Bittel, P., Chinchilla, D., Jehle, A.K., Albert, M., Boller, T., and Felix, G. (2012). Chimeric FLS2 Receptors Reveal the Basis for Differential Flagellin Perception in Arabidopsis and Tomato. The Plant Cell *24*, 2213–2224.

Naito, K., Taguchi, F., Suzuki, T., Inagaki, Y., Toyoda, K., Shiraishi, T., and Ichinose, Y. (2008). Amino Acid Sequence of Bacterial Microbe-Associated Molecular Pattern flg22 Is Required for Virulence. International Society for Molecular Plant-Microbe Interactions *21*, 1165–1174.

Nekrasov, V., Li, J., Batoux, M., Roux, M., Chu, Z.-H., Lacombe, S., Rougon, A., Bittel, P., Kiss-Papp, M., Chinchilla, D., et al. (2009). Control of the pattern-recognition receptor EFR by an ER protein complex in plant immunity. The EMBO Journal *28*, 3428–3438.

Ohno, S. (1979). Evolution by gene duplication (London: George Allen & Unwin).

Poncini, L., Wyrsch, I., Tendon, V.D., Vorley, T., Boller, T., Geldner, N., Métraux, J.-P., and Lehmann, S. (2017). In roots of Arabidopsis thaliana, the damage-associated molecular pattern AtPep1 is a stronger elicitor of immune signalling than flg22 or the chitin heptamer. PLoS ONE *12*.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLoS ONE *5*.

Rich-Griffin, C., Eichmann, R., Reitz, M.U., Hermann, S., Woolley-Allen, K., Brown, P.E., Wiwatdirekkul, K., Esteban, E., Pasha, A., Kogel, K.-H., et al. (2020). Regulation of Cell Type-Specific Immunity Networks in Arabidopsis Roots. The Plant Cell *32*, 2742–2762.

Ronald, P.C., and Beutler, B. (2010). Plant and animal sensors of conserved microbial signatures. Science *330*, 1061–1064.

Schwessinger, B., Roux, M., Kadota, Y., Ntoukakis, V., Sklenar, J., Jones, A., and Zipfel, C. (2011). Phosphorylation-Dependent Differential Regulation of Plant Growth, Cell Death, and Innate Immunity by the Regulatory Receptor-Like Kinase BAK1. PLoS Genetics *7*, e1002046.

Shinya, T., Yamaguchi, K., Desaki, Y., Yamada, K., Narisawa, T., Kobayashi, Y., Maeda, K., Suzuki, M., Tanimoto, T., Takeda, J., et al. (2014). Selective regulation of the chitin-induced defense response by the Arabidopsis receptor-like cytoplasmic kinase PBL27. The Plant Journal *79*, 56–66.

Smakowska-Luzan, E., Mott, G.A., Parys, K., Stegmann, M., Howton, T.C., Layeghifard, M., Neuhold, J., Lehner, A., Kong, J., Grünwald, K., et al. (2018). An extracellular network of Arabidopsis leucine-rich repeat receptor kinases. Nature *553*, 342–346.

Steinbrenner, A.D. (2020). The evolving landscape of cell surface pattern recognition across plant immune networks. Current Opinion in Plant Biology *56*, 135–146.

Sun, Y., Li, L., Macho, A.P., Han, Z., Hu, Z., Zipfel, C., Zhou, J.-M., and Chai, J. (2013). Structural Basis for flg22-Induced Activation of the Arabidopsis FLS2-BAK1 Immune Complex. Science *342*, 624–628.

Teixeira, P.J.P.L., Colaianni, N.R., Fitzpatrick, C.R., and Dangl, J.L. (2019). Beyond pathogens: microbiota interactions with the plant immune system. Current Opinion in Microbiology *49*, 7–17.

Thomsen, M.C.F., and Nielsen, M. (2012). Seq2Logo: a method for construction and visualization of amino acid binding motifs and sequence profiles including sequence weighting, pseudo counts and two-sided representation of amino acid enrichment and depletion. Nucleic Acids Research *40,* W281-W287.

Tran, T.M., Ma, Z., Triebl, A., Nath, S., Cheng, Y., Gong, B.-Q., Han, X., Wang, J., Li, J.-F., Wenk, M.R., et al. (2020). The bacterial quorum sensing signal DSF hijacks Arabidopsis thaliana sterol biosynthesis to suppress plant innate immunity. BioRxiv. https://doi.org/10.1101/2020.01.30.927731

Vetter, M., Karasov, T.L., and Bergelson, J. (2016). Differentiation between MAMP Triggered Defenses in Arabidopsis thaliana. PLOS Genetics *12*.

Wyrsch, I., Domínguez-Ferreras, A., Geldner, N., and Boller, T. (2015). Tissue-specific FLAGELLIN-SENSING 2 (FLS2) expression in roots restores immune responses in Arabidopsis fls2 mutants. New Phytologist *206*, 774–784.

Zhou, F., Emonet, A., Tendon, V.D., Marhavy, P., Wu, D., Lahaye, T., and Geldner, N. (2020). Co-incidence of Damage and Microbial Patterns Controls Localized Immune Responses in Roots. Cell *180*.

Zhou, J.M., and Zhang, Y. (2020). Plant Immunity: Danger Perception and Signaling. Cell *181*, 978–989.

Zipfel, C., Kunze, G., Chinchilla, D., Caniard, A., Jones, J.D., Boller, T., and Felix, G. (2006). Perception of the Bacterial PAMP EF-Tu by the Receptor EFR Restricts Agrobacterium-Mediated Transformation. Cell *125*, 749–760.

CHAPTER 4: SPECIFIC MODULATION OF THE ROOT IMMUNE SYSTEM BY A COMMUNITY OF COMMENSAL BACTERIA[4]

Section 4.1: Introduction

Plants are inhabited by hundreds of species of commensals, many of which have beneficial effects on the host (1). These microbes often express the same immunogenic MAMPs that are found in pathogens, highlighting their potential to trigger immune responses in their hosts (2–4). How plants mount effective defenses against pathogens while allowing the colonization of commensals remains a mystery. Plant-associated microbial communities are much less diverse than those of the surrounding environment (1), indicating that the host exerts selection pressure over their microbiota and that some microbes are better adapted to colonize plant tissues than others. While multiple environmental and genetic factors likely orchestrate microbiota assembly and structure, recent research indicates that the plant immune system operates as a major gatekeeper. *Arabidopsis thaliana* plants (hereafter Arabidopsis) compromised in the signaling of the defense phytohormones salicylic acid and jasmonic acid harbor altered microbiota (5, 6). Similarly, mutants impaired in MTI and in the MIN7-vesicle trafficking pathway carry altered endophytic phyllosphere microbiota and display leaf-tissue damage associated with dysbiosis

[4] This chapter is part of a manuscript published in PNAS. The original citation is as follows: Teixeira, P. J. P. L.*, **Colaianni, N*,** Law, T.F.*, Conway, J.M.*, Gilbert, S., Li, H., Panda, D., Del Risco, N.M., Finkel, O.M., Castrillo, G., Mieczkowski, P., Jones, C.D., Dangl, J.L. Specific modulation of the root immune system by a community of commensal bacteria, *PNAS,* 2021, 10.1073/pnas.2100678118**.** Note that * indicates that these authors contributed equally to the work. My contributions to this work included statistical analyses, RNA-seq and 16S analyses, modeling, identification of secretion systems, and figure creation. I also intellectually helped in experimental design and overall data interpretation.

under conditions of high humidity (7). Recent evidence demonstrates that perception of flg22 is usually low in most root cells, but up-regulated following tissue wounding associated with infection and potentially colonization (8, 9).

The ability to suppress the host immune response is a hallmark of successful pathogens (10). In animals, both specific and redundant immunomodulatory effects have been defined for taxonomically diverse commensal gut bacteria (11). Likewise, plant-associated commensals have been shown to modulate MTI (12–19). Nevertheless, studies in plants have so far focused on single microbes and the specific immunomodulatory effects of different strains have not been integrated in the context of complex communities. Furthermore, the significance of immunomodulation for community assembly remains unexplored. In this work, we investigated how a community of root-associated commensal bacteria interacts with the Arabidopsis immune system. We verified that the ability to suppress MTI is common and taxonomically widespread among these commensals. High-throughput gene expression analyses of plants colonized by synthetic communities (SynComs) or by single microbes led to the identification of a set of defense-related genes that were commonly manipulated by phylogenetically distinct suppressor strains, highlighting a sector of the plant immune system that may control plant colonization by the microbiota. Notably, suppressors could promote the growth of other non-suppressor strains, indicating that certain microbes may benefit from co-occurring suppressor strains. Furthermore, a genetic screening revealed that the type-2 secretion system is required for the robust suppressor *Dyella japonica* MF79 to interfere with root MTI, while the type-3 secretion system was dispensable. Our results expand our understanding of how the plant immune system functions in roots colonized by commensals, underscoring the role of MTI and of its suppression in microbiota assembly.

Section 4.2: Results

4.2.1: A community of root commensals suppresses MTI

To understand how MAMP-triggered immunity affects the composition of the root microbiome and whether root-associated bacteria can modulate this immune response in the context of a commensal community, we established a model system (see SI Appendix, Fig. 4.6*A*-*E*). Here, 7-day-old Arabidopsis seedlings were exposed for twelve days to 1 μM of the bacterial peptide MAMP flg22. Although MTI has been traditionally studied within minutes/hours of MAMP exposure (15, 20, 21), we validated that the continuous presence of flg22 in our system induces a defense-related transcriptional signature that is functionally overlapping with the one observed in 'typical' systems (SI Appendix, Fig. 4.6*F*), indicating that flg22 signaling and immunogenicity persist for at least twelve days. Using this system, we concurrently inoculated plants with 1 μM flg22 or not and a 35-member bacterial synthetic community composed of a random consortium of Arabidopsis root commensals (SynCom35; SI Appendix, Dataset S1) that is phylogenetically representative of wild-soil root associated microbiota (22). Interestingly, bacterial 16S profiles revealed only minor differences between the communities of bacteria colonizing flg22-treated roots and those colonizing untreated plants (Fig. 4.1*A*). Only two bacterial taxa had differential abundances in the roots of plants exposed to flg22 (SI Appendix, Fig. 4.6*G* and Dataset S2). Thus, overall community composition was not affected by the addition of the elicitor flg22, suggesting that SynCom35 interferes with the plant immune response to facilitate community assembly in the face of flg22-driven MTI.

To investigate whether SynCom35 interferes with the flg22-induced immune response in roots, we cultivated plants in axenic conditions, in the presence of SynCom35, or in the presence of heat-killed SynCom35 (2 hours at 100ºC; HK; SI Appendix, Fig. 4.6*C*-*E*). Roots exposed to

flg22 for twelve days without any bacteria or in the presence of the heat-killed SynCom35 displayed a clear transcriptional reprogramming, with 492 and 523 genes responding to the elicitor respectively (Fig. 4.1*B* and SI Appendix, Dataset S3). The flg22-responsive genes from these control treatments were not only highly overlapping (Fig. r.1*C*), but also showed highly correlated expression levels (Fig. 4.1*D*), demonstrating that the dead bacterial community does not interfere with the plant response to flg22. In contrast, only 7 genes responded to flg22 in plants inoculated with the live SynCom35 (Fig. 4.1*B*), and no correlation was observed between this condition and the two other controls (Fig. 4.1*C*-*D*). Indeed, SynCom35 triggered a unique transcriptional signature that persisted in the presence of flg22 (Fig. 4.1*E*). Hierarchical clustering revealed a set of 281 genes (39% of the total flg22-responsive genes) that were clearly activated by flg22 in the roots of control plants, but strongly suppressed when SynCom35 was also present (cluster SC1; Fig. 4.1*F*-*G*). Cluster SC1 is strongly enriched in immune-related genes (Fig. 4.1*H*), demonstrating that this bacterial commensal community possesses immunomodulatory activity.

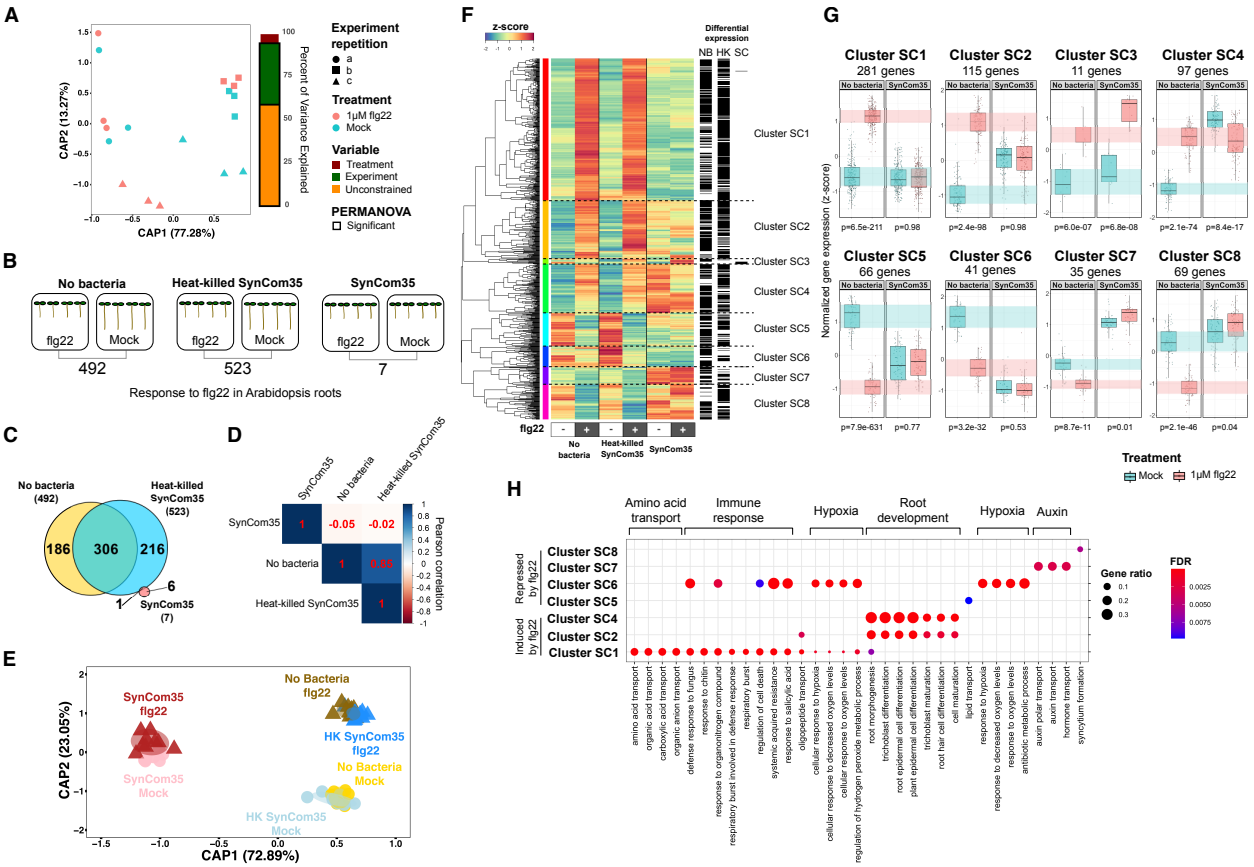


Figure 4.1. A synthetic community comprised of 35 bacterial strains (SynCom35) modulates a sector of the Arabidopsis immune system. (A) Canonical analysis of principal coordinates (CAP) based on Bray-Curtis dissimilarities of 16S rDNA profiling reveals no significant effect of flg22 in the composition of root bacterial community. Significance of model terms was determined with a PERMANOVA analysis using 5000 permutations. (B) Number of differentially expressed genes (DEG) identified by RNA-seq in roots treated with flg22 or not in the absence of bacteria, with heat-killed SynCom35 or with SynCom35 alive. (C) Overlap among the sets of flg22-responsive genes in the 'No bacteria', 'Heat-killed SynCom35' and 'SynCom35' conditions. (D) Pearson correlation among flg22 responses in plants grown axenically (no bacteria) or in the presence of dead or living SynCom35. Note the low correlation between 'SynCom35' and the other two conditions. (E) Canonical Analysis of Principal coordinates shows that plants grown in the presence of SynCom35 display a unique transcriptional signature that is not differentiated by addition of flg22. (F) Hierarchical clustering of the 716 genes that were differentially expressed by exposure to flg22 in at least one of the experimental conditions. SynCom35 triggers a transcriptional signature that is very similar to the flg22 response except for three sets of genes (Cluster SC1, Cluster SC7 and Cluster SC8), which display a remarkably different expression profile when compared to control plants. (G) Expression pattern of each of the clusters defined in the heatmap. SynCom35 inhibits the flg22 response of clusters SC1, SC2, SC5, SC7 and SC8. The p-values on the bottom refer to the comparison between groups (flgs22 vs mock) for each condition using t-tests. (H) Gene ontology enrichment analyses showing enriched biological processes in each of the clusters. Cluster SC1, which is suppressed by SynCom35, contain defense genes that are activated by flg22. Cluster SC7, which is activated by SynCom35, is enriched in auxin metabolism. All results refer to plants exposed to the treatments for 12 days. NB: No bacteria, HK:

Heat-killed SynCom35; SC: SynCom35. The complete enrichment analysis and the associated statistics are shown in SI Appendix, Dataset S3

WRKY transcription factors, which include master regulators of plant immunity (23), were overrepresented in cluster SC1 (p = 8.8E-6; SI Appendix, Fig. 4.7*A*-*E*), which contained 7 of the 8 *WRKY* genes that were up-regulated by flg22. Consistent with a suppressed WRKY regulatory network, cluster SC1 was also enriched in predicted targets of these transcription factors (SI Appendix, Fig. 4.9*F*). SynCom35 also suppressed the activation of genes encoding Receptor-Like Kinases (RLKs; SI Appendix, Fig. 4.9*G*), including the MTI marker *FRK1* (FLG22-INDUCED RECEPTOR-LIKE KINASE 1) and the flg22 receptor itself, *FLS2* (FLAGELLIN-SENSITIVE 2). RLKs are receptors that form the first line of defense in the plant immune system by transducing the perception of extracellular molecules into intracellular signals that activate immunity (24). Genes involved in the biosynthesis of secondary metabolites were also prevalent in cluster SC1. In particular, the master regulator of indole glucosinolate biosynthesis *MYB51* (25), along with three indole glucosinolate methyltransferases genes (*IGMT2*, *IGMT3* and *IGMT4*) and various glutathione S-transferases and cytochrome P450s were not activated by flg22 in plants colonized by SynCom35 (SI Appendix, Fig. 4.7*H*). Root-exuded secondary metabolites act as chemical barriers against microbial invaders and suppression of their biosynthesis may favor colonization of microbial commensals (26). These results demonstrate that key components of the plant immune system, from receptors to transcription factors and biochemical executors, are modulated by SynCom35.

Since cluster SC1 overlaps significantly with a set of flg22-responsive genes that were suppressed by a different and independently derived SynCom from an independent study (SI Appendix, Fig. 4.8*A* and Dataset S3) (27), modulation of this sector of the plant immune system is likely to be a common strategy employed by root commensals. Cluster SC1 also overlaps

significantly (84/281 genes; p-value=3.27E-58; SI Appendix, Dataset S3) with a set of 868 genes that are flg22-regulated across four *Brassicaceae* species (28), indicating that SynCom35 has the potential to modulate an evolutionarily conserved sector of MTI. Although defense-related genes were not activated by flg22 in plants grown with SynCom35 at day 12, they were up-regulated in the first day of treatment (SI Appendix, Fig. 4.8*B*-*F* and Dataset S3), indicating that SynCom35, once fully established, suppresses an ongoing immune response instead of preventing its activation.

The subset of flg22-activated genes that were not suppressed by SynCom35 at day 12 (clusters SC2 and 4; Fig. 4.1*F*) were enriched in root morphogenesis and differentiation processes (Fig. 4.1*H*). These genes were constitutively activated by SynCom35 even in the absence of flg22 (Fig. 4.1*G*). Impaired primary root elongation is a phenotype of plants exposed to flg22 (29), a phenotype that is also triggered by SynCom35 (SI Appendix, Fig. 4.9*A*). *Arthrobacter* strains present in SynCom35 affect root development, likely via auxin production (30). Consistent with this observation, auxin-responsive genes were highly activated in plants grown with SynCom35 (cluster SC7; Fig. 4.1*F*-*H*). Furthermore, the auxin-resistant mutant *axr2-1* expressed a significantly reduced root growth-inhibition (RGI) phenotype in the presence of *Arthrobacter* MF161 (SI Appendix, Fig. 4.9*B*). These results demonstrate that the SynCom35 effect on root morphology is uncoupled from its immunomodulatory activity.

4.2.2: Dissecting the immunomodulatory activity of commensals

After establishing that SynCom35 had immunomodulatory activity on Arabidopsis roots, we performed mono-association experiments to evaluate the interaction of individual strains with the plant immune system. Seven-day-old Arabidopsis seedlings were inoculated with each member of SynCom35 in the presence or absence of 1 μM flg22 (SI Appendix, Fig. 4.10).

Evaluation of root lengths revealed that individual strains interacted with the plants in three different ways (Fig. 4.2*A*). Twenty strains did not alter the flg22 effect on root morphology. Seven strains caused root-growth inhibition (RGI) even in the absence of flg22 (RGI inducers), a phenotype reminiscent of the SynCom35 effect on the plant. Importantly, eight strains suppressed the RGI phenotype triggered by flg22 (RGI suppressors), supporting longer root lengths in the presence of the MAMP. Because plants grown in the presence of SynCom35 display short roots (Fig. 4.2*A* and SI Appendix, Fig. 4.10*A*), the RGI inducers in this consortium are functionally dominant over the suppressors for this trait.

We next asked whether strains that suppressed flg22-mediated RGI were also efficient suppressors of the flg22 regulon. A Principal Component Analysis (PCA) based on the expression of 428 genes that define the flg22 regulon in axenic plants (SI Appendix, Dataset S5) revealed suppression of this transcriptional response by various members of SynCom35 (Fig. 4.2*B*), including all strains that suppressed flg22-mediated RGI (Fig. 4.2*A*). Thus, efficient suppression of the flg22 regulon predicts suppression of the flg22 effect on root elongation. This conclusion was validated using an orthogonal approach. We generated a literature-based core set of 84 flg22 transcriptional markers, defined by both chronic and acute flg22 treatments of diverse Arabidopsis tissues (SI Appendix, Fig. 4.11*A*-*B* and Dataset S5). These 84 genes contained 40 cluster SC1 genes (p=2.82E-57) and 52 of the evolutionarily conserved flg22 regulon (28) (p=9.34e-56). We clustered our data based on their expression (SI Appendix, Fig. 4.11*C*). The two resulting groups reflected the expression levels of these defense-related genes: Group I (SI Appendix, Fig. 4.11*C*, blue) displayed low expression of these marker genes and was mostly comprised of mock-treated plants; and Group II (SI Appendix, Fig. 4.11*C*, red) presented higher expression of the same genes and was mostly comprised of flg22-treated plants. However,

some strains prevented the activation of the flg22 core set when plants were exposed to the elicitor, while others activated the same genes even in the absence of flg22. These results show that the SynCom35 commensal consortium contains members that have opposing effects on the plant immune response to flg22.

To explore strain-specific effects on the plant immune response in more detail, the 428 flg22-responsive genes that were defined in axenic plants were subjected to hierarchical clustering based on their expression levels in the context of mono-associations (Fig. 4.2*C*). Eleven of the 35 strains (31%) had no effect on the plant response to flg22; seedlings grown in the presence of these bacteria responded to the MAMP similarly to seedlings grown without bacteria or with the heat-killed SynCom35 (Fig. 4.2*C*). The remaining 24 strains (69%) suppressed the plant transcriptional response to flg22 either partially or completely. Of the three clusters containing genes activated by flg22 (clusters M3, M4 and M5; Fig. 4.2*C*), one was enriched in root development (cluster M3; Fig. 4.2*C* and SI Appendix, Dataset S4) and two were enriched in defense-related genes (clusters M4 and M5; Fig. 4.2*C* and SI Appendix, Dataset S4). Importantly, cluster M5 was the only set of flg22-responsive genes whose activation by the MAMP was consistently suppressed by all 24 strains, indicating that modulation of these defense-related genes is a common feature shared among phylogenetically diverse commensal bacteria. Cluster M5 is also highly overlapping (76 out of 120 genes) with the cluster SC1 genes suppressed by SynCom35 in Arabidopsis roots (Fig. 4.2*C*, column labeled Roots SC1; Fig. 4.1*F*).

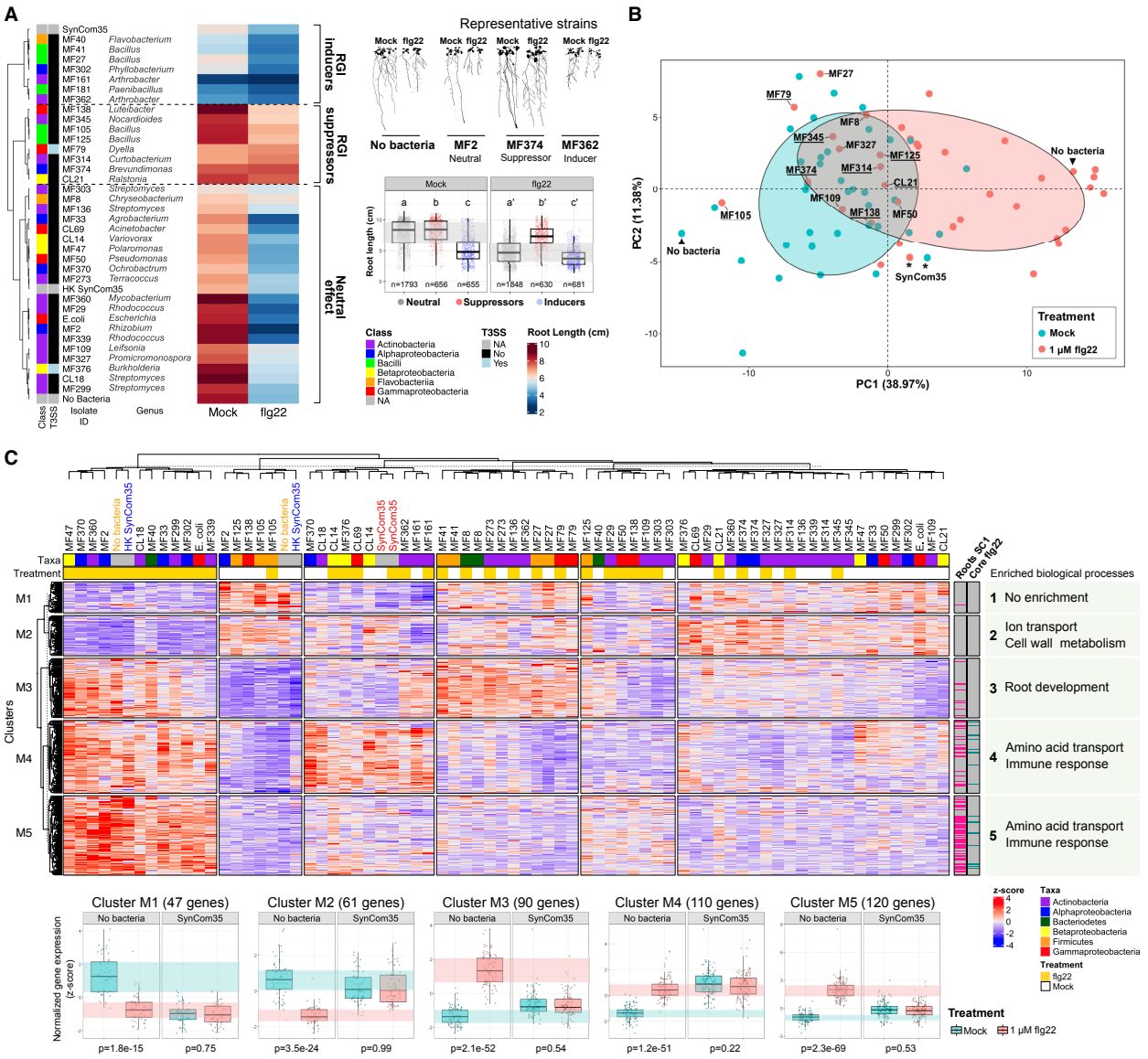


Figure 4.2. Suppression of the flg22 response is a common feature among members of SynCom35. (A) Arabidopsis root length in the presence of each strain of SynCom35. While some strains have a neutral effect on the plant root elongation, others are able to suppress the flg22-induced root growth inhibition (RGI) or induce the same phenotype even in the absence of flg22. The taxonomic class of each strain and the presence or not of a type-3 secretion system (T3SS) are annotated as colored boxes on the left of the heatmap. The images on the right panel show the plant phenotype in the presence of a representative neutral (MF2), suppressor (MF374) and inducer strain (MF362). The boxplots summarize the effect of each group from the heatmap on root length (multiple comparisons were performed with ANOVA followed by a Tukey test, $\alpha = 0.05$). (B) Principal component analysis based on 428 genes that responded to flg22 in axenic plants demonstrates that some, but not all, strains of SynCom35 interfere with the transcriptional regulation of the flg22 regulon. Ellipses show the parametric smallest area around the mean that contains 70% of the probability mass for each group (blue: mock; red: 1 µM flg22). Only controls ("no bacteria" and "SynCom35") and suppressors of this gene set are labeled for clarity. Underlined labels indicate that the strain also suppressed the RGI phenotype triggered by flg22

shown in (A). (C) Hierarchical clustering based on the expression of the same 428 genes in seedlings grown in the presence of each strain from SynCom35 along with controls (no bacteria, heat-killed SynCom35 and SynCom35). Five clusters were defined and their profiles in seedlings grown under axenic conditions or with SynCom35 are shown on the bottom. The p-values refer to the comparison between groups (flg22 vs mock) using t-tests. The boxes on the top of the heatmap indicate the taxonomic classification of the inoculated strain and treatment (orange) or not (white) with flg22. Lanes on the right of the heatmap indicate whether each gene was suppressed by SynCom35 in roots (cluster SC1, Figure 4.1*F*) or whether they belong to the set of 84 core flg22 marker genes (SI Appendix, Fig. 4.11). Representative biological processes that are enriched in each cluster are shown on the right. The complete enrichment analysis and the associated statistics are shown in SI Appendix, Dataset S4.

Evaluation of the Arabidopsis transcriptome in response to each member of SynCom35 in the absence of exogenous MAMPs complemented the analyses focused on the flg22 response (SI Appendix, Fig. 4.12*A-B*). Most strains (22/35) triggered the activation of defense-related genes to some extent (SI Appendix, Fig. 4.12*C*). However, activation of defense-related genes was undetectable in plants grown in the presence of the other, taxonomically diverse, 13 strains (SI Appendix, Fig. 4.12*C*). Eleven of these strains also suppressed cluster M5 of flg22-responsive genes in plants treated with flg22 (Fig. 4.2*C*). Thus, these results extend our observations and demonstrate that phylogenetically diverse commensals can modulate immune responses triggered by endogenous MAMPs.

We assessed the modulation of MTI by root commensals using an independent and conceptually distinct experimental system. Because flg22 triggers transcriptional responses in plant cells within minutes and over hours of its perception (15, 20, 21), we evaluated whether members of SynCom35 also interfere with the acute response that is triggered only 5 hours after root exposure to flg22 using the same setup that was previously employed to characterize MTI in the roots (19). For this, we utilized an Arabidopsis line carrying the *GUS* reporter gene under control of the flg22-responsive promoter *pCYP71A12*, which is activated by flg22 specifically in the root elongation zone (19). Plants treated with 100 nM flg22 under axenic conditions activated

the *pCYP71A12* promoter in roots within 5 hours of treatment. This response was absent in the control treatments (i.e., plants treated with 10 μM MeJA or with the *Pseudomonas simiae* strain WCS417 (19)) but was observed in plants inoculated with most of the members of SynCom35 (SI Appendix, Fig. 4.11*D*). Importantly, seven strains suppressed *pCYP71A12* activation by flg22 in this assay, indicating that these commensals can also interfere with the plant acute response to this MAMP.

By collating the results from the independent and conceptually different mono-association experiments, a set of ten robust MTI suppressors emerged within the members of SynCom35 (Fig. 4.3*A*). Although some strains of this community remained neutral or even activated the plant immune system in some assays, a total of 23 strains (65%) suppressed the plant response to flg22 in at least one assay (Fig. 4.3*A*). These results confirm that the ability to suppress the plant immune system is common and taxonomically widespread among root commensals. Interestingly, no strain among six tested robust suppressors modulated root immunity by lowering the extracellular pH (SI Appendix, Fig. 4.13*A*), a strategy recently described for commensal *Pseudomonas* (12). Interestingly, some suppressor strains interfered with the plant response to other peptide ligands and prevented RGI phenotypes induced by the damage-associated pep1 and developmental regulator clv3 peptides, while failing to suppress the RGI phenotype induced by hormones methyl jasmonate or auxin (SI Appendix, Fig. 4.13*B*). Thus, a diversity of extracellular MTI-suppression molecular mechanisms likely evolved among plant-associated bacteria in order to interfere with plant immunity.

4.2.3: MTI can control the colonization of commensals

We then tested the hypothesis that the ability to suppress the plant immune response is advantageous to commensal bacteria and, therefore, has ecological and evolutionary

significance. We first evaluated the colonization ability of five robust suppressors and taxonomically matched strains from SynCom35 that did not suppress all measured outputs (Fig 4.3*A*; SI Appendix, Dataset S1). Mono-association assays demonstrated that most suppressors grew to higher titers in Arabidopsis roots than non-suppressors (Fig. 4.3*B*). This conclusion was supported by experiments using two smaller SynComs made of the five suppressors (SynCom5-S1) and five taxonomically matched non-suppressors (SynCom5-NS1). In agreement with the mono-association assays, 10-fold more bacterial colonies were recovered from Arabidopsis roots inoculated with SynCom5-S1 than with SynCom5-NS1 (Fig. 4.3*C*). The same result was obtained in an independent experiment that used different SynComs (SynCom5-S2 and SynCom5-NS2; SI Appendix, Dataset S1 and SI Appendix, Fig. 4.14). As predicted, seedlings inoculated with SynCom5-NS1 showed stronger activation of defense-related genes in comparison to seedlings inoculated with SynCom5-S1 (Fig. 4.3*D*; SI Appendix, Fig. 4.14*B*-*E*; SI Appendix, Dataset S6). These genes encompass well-characterized components of the plant immune system, including those involved in the biosynthesis of glucosinolates, as well as RLKs and WRKYs (SI Appendix, Fig. 4.14*D*). Many of these genes were also found in clusters SC1 (Fig. 4.1*F*) and M5 (Fig. 4.2*C*). Thus, efficient suppression of MTI by commensals correlates with enhanced root colonization. Supporting this conclusion, suppressors were detected at higher relative abundances more often than non-suppressors in roots colonized with SynCom35 (Fig. 4.3*E*). Furthermore, strains that we defined as robust suppressors were also classified as robust endophytic compartment colonizers in a previous study using SynCom35 (22), while strains that we classified as non-suppressors were not (Fig. 4.3*B*).

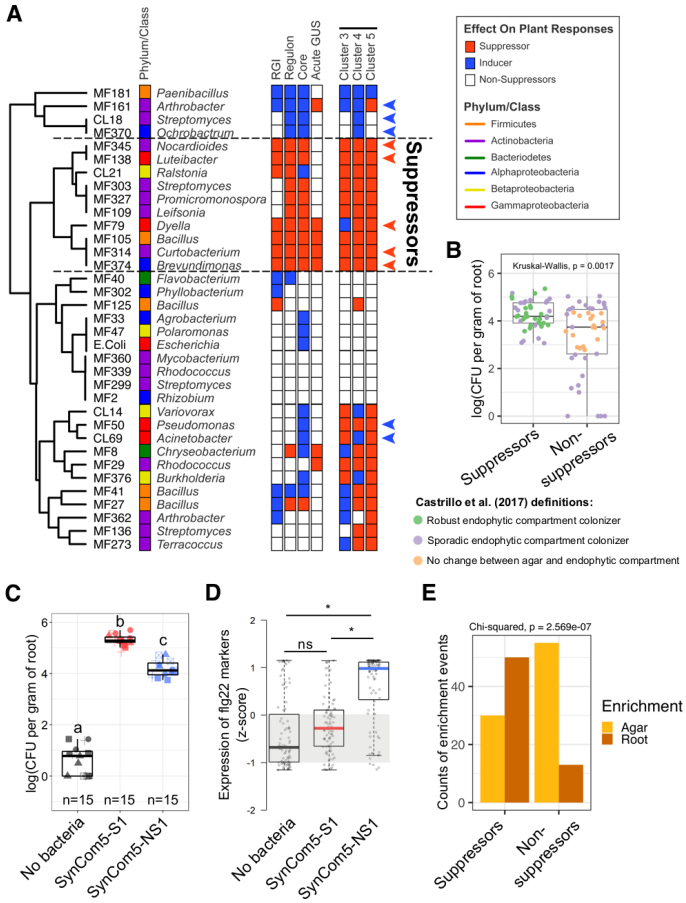


Figure 4.3. Microbiota members that suppress the plant immune system are better colonizers of Arabidopsis roots either in mono-association or in the context of SynComs. (A) Summary of the assays performed to define suppression of the flg22 response in this study. Strains are clustered based on their response to a total of 5 assays that are illustrated as rows of boxes to the right. RGI: the ability of a strain to suppress flg22-induced root growth inhibition or induce this phenotype even in the absence of the MAMP (Fig. 2*A*). Regulon: the ability to interfere with the overall expression signature of 428 flg22-responsive genes defined in control (no bacteria) seedlings (Fig. 2*B*). Core: the ability of each strain to interfere specifically with a set of 84 defense-related genes that are consistently induced by flg22 under diverse experimental conditions (SI Appendix, Fig. S6*C*). Acute GUS: the ability of each strain to interfere with the acute response to flg22 (5h treatment) specifically in roots based on the *pCYP71A12::GUS* reporter line (SI Appendix, Fig. 4.11*D*). Clusters 3, 4 and 5: the suppression ability of each strain on specific transcriptional response sectors of the flg22 regulon (Fig. 4.2*C*). Members of SynCom5-S1 and SynCom5-NS1 are marked with red and blue arrows, respectively. (B) Suppressors are better colonizers of Arabidopsis roots in mono-association experiments. The growth of five suppressors (members of SynCom5-S1) and taxonomically matching non-suppressors (members of SynCom5-NS1) were individually assayed in Arabidopsis roots after 12 days of colonization in three biological replicates. Each dot represents an individual strain/replicate. Strains are colored based on their behavior in the context of SynCom35 in a previous study (22). Note that all robust endophytic compartment colonizers are suppressor strains. A Kruskal-Wallis test was performed on the CFU counts from 3 independent experiments each with 3 technical reps. n=9 for each strain and n=45 for each group being compared. (C) A SynCom made of five suppressors (SynCom5-

S1) colonizes Arabidopsis roots better than a SynCom made of five taxonomically matching non-suppressors (SynCom5-NS1). Different symbols represent independent experimental repetitions. Letters indicate significantly different groups based on an ANOVA followed by a Tukey test (α = 0.05). (D) Seedlings grown with SynCom5-NS1 activate the expression of flg22 marker genes while seedlings inoculated with SynCom5-S1 do not. Statistical significance was determined using a permutation approach, where the actual group mean differences were compared to the group mean differences of 10,000 permutations of the gene counts. The group mean differences that were greater than 95% of the group mean differences calculated in the permutations were considered significant (*). (E) Suppressor (SynCom5-S1), but not non-suppressor (SynCom5-NS1), strains are enriched in the endophytic compartment of Arabidopsis roots relative to agar in the context of a synthetic community. Suppressors n=80, non-suppressors n=68.

Since suppressor and non-suppressor strains co-occur in the root microbiota, we evaluated whether modulation of MTI by suppressors enhances the colonization ability of other commensal strains (Fig. 4.4*A*). For this, Arabidopsis seedlings were initially grown for 14 days in the presence of synthetic communities comprised of five suppressors (SynCom5-S2) or five non-suppressors (SynCom5-NS2) (SI Appendix, Dataset S1). Subsequently, plants were flood-inoculated for 5 min with the inducer *Ochrobactrum sp.* MF370 and grown for 48 hours. More *Ochrobactrum sp.* MF370 cells were recovered from roots colonized with SynCom5-S2 compared to those inoculated with SynCom5-NS2 or grown axenically (Fig. 4.4*B*). The same result was observed for a second commensal microbe, *Pseudomonas viridiflava* OTU5 strain p5.e6 (Fig. 4.4*C*), which can also be an opportunistic plant pathogen (31). Thus, suppressor strains can enhance the root colonization capacities of other microbes, indicating that MTI, and its suppression, can contribute to the control of microbial loads in the roots.

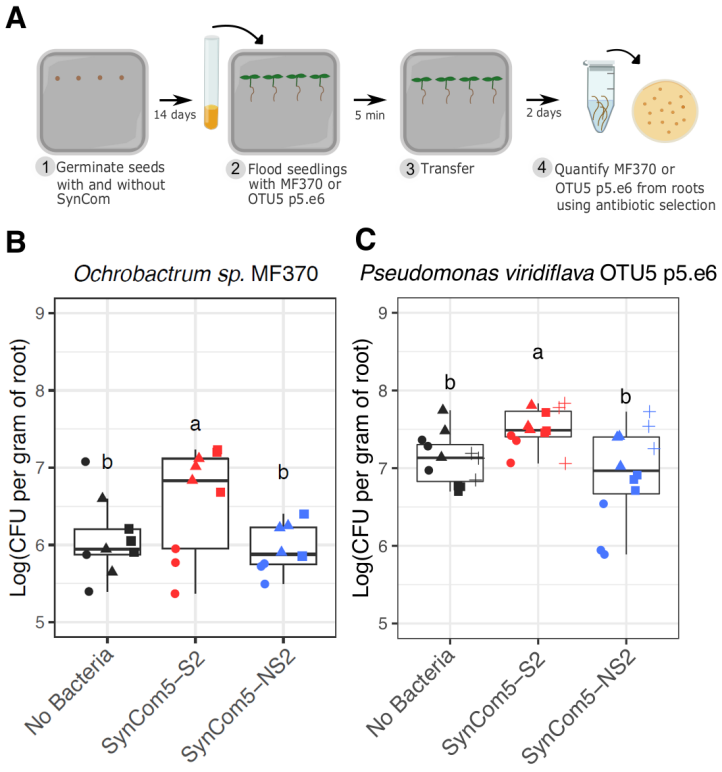


Figure 4.4. Immunomodulatory bacteria enhance the root colonization ability of other commensal strains. (A) Cartoon representation of the experimental system employed. Arabidopsis seedlings were grown axenically, in the presence of a community of five suppressors (SynCom5-S2) or in the presence of five non-suppressors (SynCom5-NS2) and then inoculated with one of the commensal strains. (B) Quantification of *Ochrobactrum sp.* MF370 cells (kanamycin resistant) from Arabidopsis roots (n=9). (C) Quantification of *Pseudomonas viridiflava* OTU5 p5.e6 cells (gentamycin resistant) from Arabidopsis roots (n=12). (B-C) Different symbols represent independent experimental repetitions. Multiple comparisons were performed with ANOVA followed by a Tukey test ($\alpha = 0.05$).

4.2.4: Immunomodulatory strategies in commensals

Finally, we sought to identify the mechanism of suppression by one of the robust suppressor strains, *Dyella japonica* MF79 (Fig. 4.3*A*). This strain carries a type-3 secretion system (T3SS), a molecular syringe that is widely employed by bacterial pathogens to inject virulence effectors into the cells of their hosts to suppress MTI (32). Thus, we tested the hypothesis that this commensal relies on the T3SS to interfere with plant immunity. We constructed two T3SS deletion strains and evaluated their ability to modulate the root response to

flg22 using the *pCYP71A12::GUS* reporter line. The first mutant (MF79 ΔT3SS_RSTU) deletes the T3SS inner membrane components (*SctRSTU*), whereas the second mutant (MF79 ΔT3SS_full) deletes the full T3SS locus (SI Appendix, Fig. 4.15*A*). Interestingly, neither of these mutants were impaired in their ability to suppress the root response to flg22 (Fig. 4.5*A*). This demonstrates that the T3SS in this strain is dispensable for immune suppression.

To identify novel *D. japonica* MF79 genes required for the suppression activity, approximately 4,500 strains from a transposon insertion library (33) were screened on the *pCYP71A12::GUS* reporter line in a high-throughput 96-well format. We found six mutant strains that were unable to suppress the activation of the *pCYP71A12::GUS* reporter by flg22 (Fig. 4.5*B*). Remarkably, all six transposon insertions mapped to the type-2 secretion system (T2SS): one to the intracellular component *gspE* and five (four independent transposon hits, one duplicate) to the outer membrane component *gspD* (SI Appendix, Fig. 4.15*B* and Dataset S7). Thus, the T2SS is required for the secretion of substrates with immune suppressive activity to the extracellular space. Wild-type and mutant strains grew at the same rate *in vitro* (SI Appendix, Fig. 4.15*C*), suggesting that the loss of suppression ability in the *gspD* and *gspE* mutants is not merely due to growth defects. This was confirmed by experiments in which supernatants of bacteria grown in plant-free medium were used. Although no bacterial growth differences were observed *in vitro*, supernatant from wild-type *D. japonica* MF79 suppressed the root response to flg22, while neither the *gspD* nor *gspE* mutant supernatants had suppressive activity (Fig. 4.5*C*). The suppressive activity from wild-type *D. japonica* MF79 was retained on the filter of a 10 kDa protein concentrator, but not in the flow-through (Fig. 4.5*D*). Importantly, the 10 kDa retentate from wild-type *D. japonica* MF79 prevented the flg22-mediated reduction in root colonization of the non-suppressor strain *Ochrobactrum sp.* MF370, and this effect required type 2 secretion

components (Fig. 4.5*E*). Both the *gspD and* the *gspE* mutants were slightly poorer colonizers of Arabidopsis roots than the wild-type strain in mono-association assays (SI Appendix, Fig. 4.15*D*). Taken together, these results suggest that suppression of the plant response to flg22 by *D. japonica* MF79 is mediated by at least one T2SS-secreted protein that is larger than 10 kDa and that the suppressive ability correlates with enhanced colonization of Arabidopsis roots.

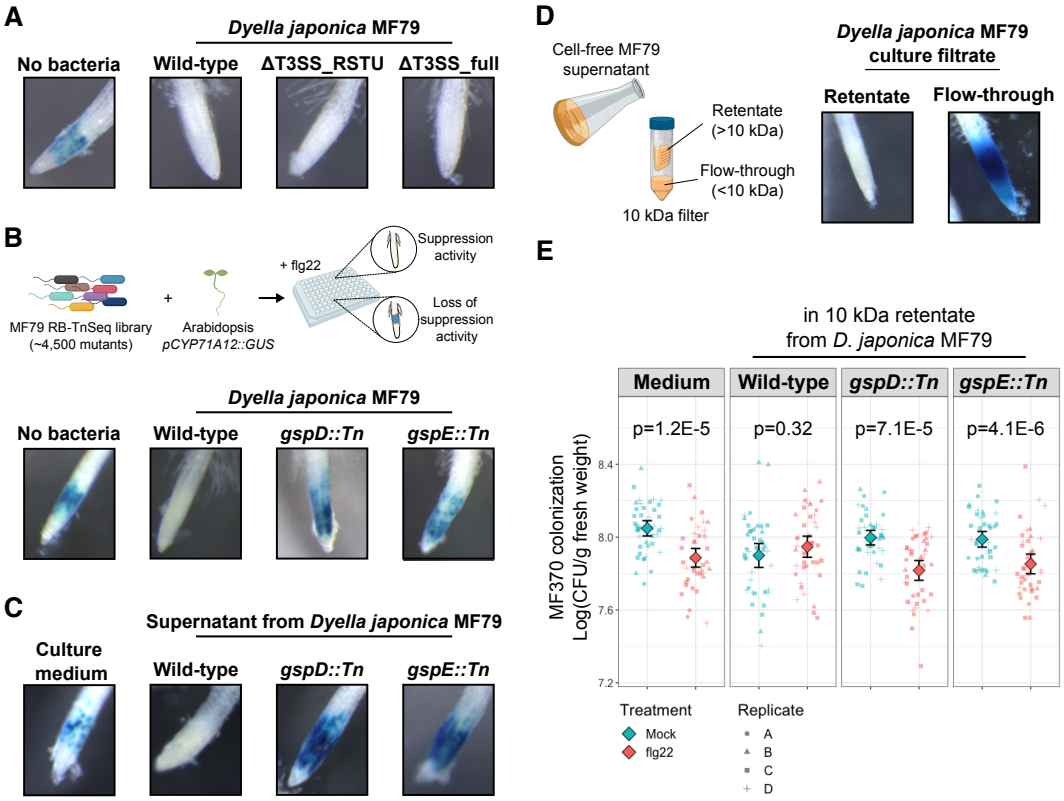


Figure 4.5. *Dyella japonica* MF79 requires the type-2 secretion system (T2SS) to suppress the root response to flg22. (A) The type-3 secretion system (T3SS) is not required for MTI suppression by *D. japonica* MF79. Two independent mutants (ΔT3SS_RSTU and ΔT3SS_full) retained the ability to prevent activation of the *pCYP71A12::GUS* reporter in roots exposed to 100 nM flg22 for 5 hours. (B) The screening of a transposon insertion library revealed two major components of the bacterial T2SS (GspD and GspE) as required for suppression of the root response to flg22. The diagram on the top summarizes the strategy employed in the screening. Additional mutants and replicates are shown in SI Appendix, Fig. 4.15*B*. (C) The cell-free supernatant of wild-type *D. japonica* MF79, but not of T2SS mutants, is sufficient for the suppression of the root response to flg22. (D) The molecule responsible for the suppression activity is larger than 10 kDa. The diagram on the left illustrates the preparation of 10 kDa retentate and flow-through from bacterial cultures. (E) The 10 kDa retentate of wild-type *D. japonica* MF79 prevents the flg22-mediated reduction of *Ochrobactrum sp.* MF370 colonization. Four independent experiments were performed (n=42 per condition). Diamonds represent means with two times standard error. p-values were determined with the Wilcoxon test.

Section 4.3: Discussion

MTI is a well-known plant immune response to invading microbes, and its modulation by disease-causing pathogens has been extensively reported (10). Recent evidence suggests that the ability to suppress MTI has also evolved in commensal microbes (2). While previous studies focused on single microbes and on relatively simple MTI readouts, we demonstrate that a complex bacterial community of Arabidopsis root commensals interferes with a specific sector of the plant transcriptional response to the MAMP flg22 (cluster SC1; Fig. 4.1*F*-*G*). The suppression of these genes thus defines the set of host immune response genes that are likely modulated to facilitate successful commensal colonization. Suppression of subsets of MAMP-responsive genes has been observed in Arabidopsis roots colonized by the single commensal strains *P. simiae* WCS417 and *Bacillus subtilis* FB17 (15, 34). Although these studies did not investigate the effect of the microbes on the immune response elicited with exogenous MAMPs, they support our conclusion that diverse commensal bacteria interfere with specific sectors of the plant immune system. Thus, similar to what has been established for pathogenic microbes, suppression of the plant immune system is a hallmark of root colonization by commensals. Importantly, our findings agree with and significantly extend the proposal that the ability to suppress the host immune system is common and taxonomically widespread among root-associated bacteria (12, 13), a concept that has also emerged for human gut commensals (11). Recent evidence suggests that root commensals may avoid host cell damage, a signal that enhances MTI in root tissues, to prevent the activation of plant immunity during colonization (8). Since we observed a clear response to flg22 in our control conditions (no bacteria and heat-killed SynCom35), suppression of defense-related genes in roots colonized by SynCom35 is not merely

due to damage avoidance. It is noteworthy that the suppression of immune responses is a dominant trait in our SynCom.

A distinctive feature of our study is the dissection of the immunomodulatory capacity of 35 individual commensal strains and the integration of this information with the outcome observed in a community context. Mono-association experiments revealed that root commensals trigger both shared and unique responses in Arabidopsis, without any obvious taxonomic signature (Fig. 4.2; SI Appendix, Fig. 4.11). Interestingly, a common set of defense-related genes was suppressed by most strains (cluster M5; Fig. 4.2*C*), indicating that phylogenetically diverse bacteria likely employ multiple mechanisms to exert a redundant immunomodulatory effect on the host. Many of these genes were also suppressed by a different bacterial community characterized in an independent study (SI Appendix, Fig. 4.8*A*) (27), supporting the hypothesis that root commensals manipulate a core set of biological processes in their hosts. Members of the WRKY family of transcription factors (e.g. *WRKY28*, *30* and *33*) were consistently activated by flg22 in the roots, but suppressed by our strains and by those reported by Ma et al. (27) (SI Appendix, Fig. 4.8*A*). These *WRKYs* might constitute key regulators of the plant immune response to commensals, possibly playing important roles in the homeostasis of immune responses during commensal community assembly. Importantly, a robust, widespread and redundant immunomodulatory capacity among commensals could buffer plant-associated communities against perturbations and maintain homeostasis when some members are lost or when external stimuli are present, as demonstrated for SynCom35 (Fig. 4.1*A*). Furthermore, the unique effect of each strain on the host may account for specific features, such as the ability to prime the plant immune system or to trigger induced systemic resistance (35).

One hypothesis supported by our findings is that the presence of suppressor strains might benefit other non-suppressors in the context of plant-associated communities. Although suppressors were often more efficient root colonizers than non-suppressor strains, the latter (including *D. japonica* MF79 T2SS mutants) could still be recovered from Arabidopsis roots, either in mono-association or SynCom experiments. This indicates that suppression of the plant immune system likely facilitates bacterial growth within plant tissues but is unlikely to be a requirement for the colonization by most commensal strains. Interestingly, root colonization of non-suppressive bacteria was enhanced by the presence of a suppressor SynCom (Fig. 4.4). We predict that this effect might involve specific, rather than general, microbial combinations. Niche occupancy and specific microbe-microbe interactions are important factors contributing to microbial colonization of plant tissues, resulting in different community contexts and complex phenotypes. It is interesting to note, however, that cell-free supernatant from wild-type *D. japonica* MF79, but not from T2SS mutants, prevented the flg22-mediated reduction in root colonization of the non-suppressor strain *Ochrobactrum sp.* MF370 (Fig. 4.5*E*). This indicates that non-suppressor bacteria can indeed benefit from the immunomodulatory activity of other community members under specific circumstances.

Our data suggest that the ability to suppress the plant immune system has evolved several independent times in plant commensals, likely resulting in numerous immunosuppressive mechanisms (2, 4). In contrast to most pathogen or mutualist microbes, root-associated commensals seem to display a much lower degree of host specialization. As a consequence, commensals may interfere with MTI mostly through nonspecific extracellular strategies. In turn, pathogens and specialized symbionts, such as nodule-forming rhizobia, often rely on highly specialized effector proteins that are injected into the host cell through the T3SS (36).

Interestingly, *D. japonica* MF79 is the only suppressor commensal in SynCom35 to carry genes for the T3SS. Nevertheless, *D. japonica* MF79 *ΔT3SS* mutants retained their ability to suppress the flg22 response in Arabidopsis roots. Our transposon library screening resulted in six mutants in the T2SS, but no hits in other genes encoding putative T2SS substrates. This suggests that multiple proteins are secreted by *D. japonica* MF79 to the extracellular space and act redundantly to suppress MTI, consistent with previous demonstrations that T2SS can be required for pathogen virulence (37). Interestingly, only three of the ten robust suppressors of the flg22 response (Fig. 4.3*A*) are predicted to carry a T2SS (SI Appendix, Dataset S1), indicating the T2SS is dispensable for the suppression activity of most of these strains. We expect that new research will reveal a large number of novel strategies employed by commensals to modulate the plant immune system. Future studies should also elucidate the role of plant genes, specifically those in cluster SC1 (Fig. 4.1*F*-*G*), in the assembly of microbial communities.

Section 4.4: Supplemental Figures

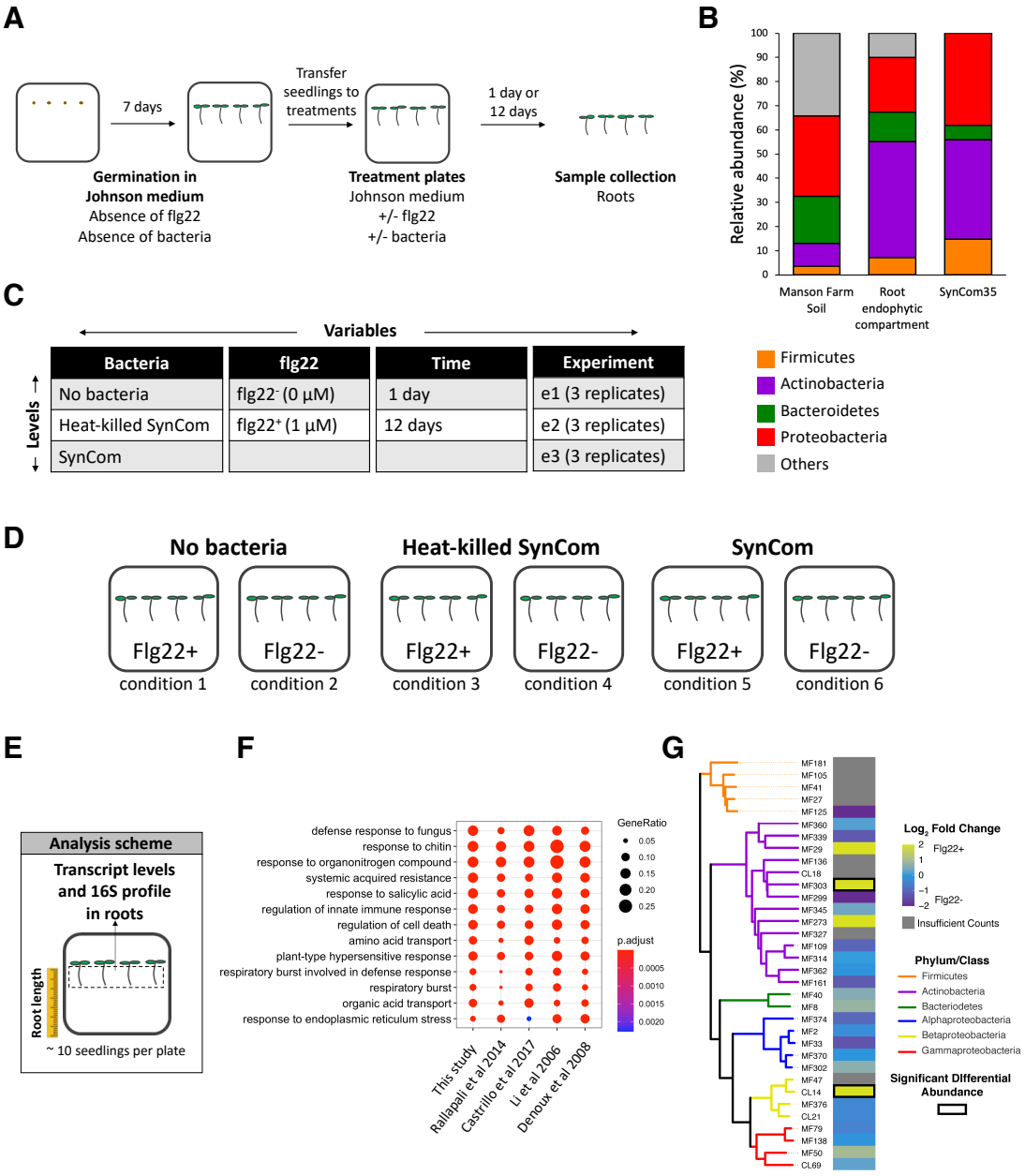


Figure 4.6. Evaluating the Arabidopsis root transcriptome and bacterial community in response to the MAMP flg22. (A) Representation of the experimental system used to evaluate the effect of SynCom35 on the root response to flg22. Seeds of wild-type Arabidopsis (Col-0 ecotype) were germinated in Johnson medium under gnotobiotic conditions in the absence of flg22. After seven days, the resulting seedlings were transferred to plates (10 per plate) containing Johnson medium embedded with SynCom35 (105 cfu/mL) and/or 1 µM flg22. Roots were harvested for RNA extraction one or twelve days later. (B) Comparison of the bacterial taxonomic composition of Manson Farm soil, the root endophytic compartment of Arabidopsis plants grown in the same soil and of SynCom35. (C) Variables accounted for in the experiment: presence of

bacteria, presence of flg22, length of the treatment and experimental batch. (D) A total of six different conditions were evaluated, corresponding to plants treated or not with 1 μM flg22 in the absence of bacteria, with heat-killed SynCom35 or with SynCom35. (E) Cartoon representation of the outputs measured: root length, root transcriptional program (RNA-seq) and root bacterial community profile (16S amplicon sequencing). (F) Seedlings treated with flg22 for 12 days activate defense-related genes in this experimental system. The figure shows biological processes (Gene Ontology) enriched among genes that were up-regulated by flg22 in control plants (in the absence of SynCom35). Four other independent studies are shown for comparison. Note that these other studies are based on plants of different ages, different tissues, different treatment times and flg22 concentration. See SI Appendix, Dataset S5 for details about each experiment. (G) Variation in relative abundance of each member that comprises SynCom35. Only two strains (MF303 and CL14) displayed significant differences in abundance in roots treated with flg22 relative to untreated roots after 12 days.

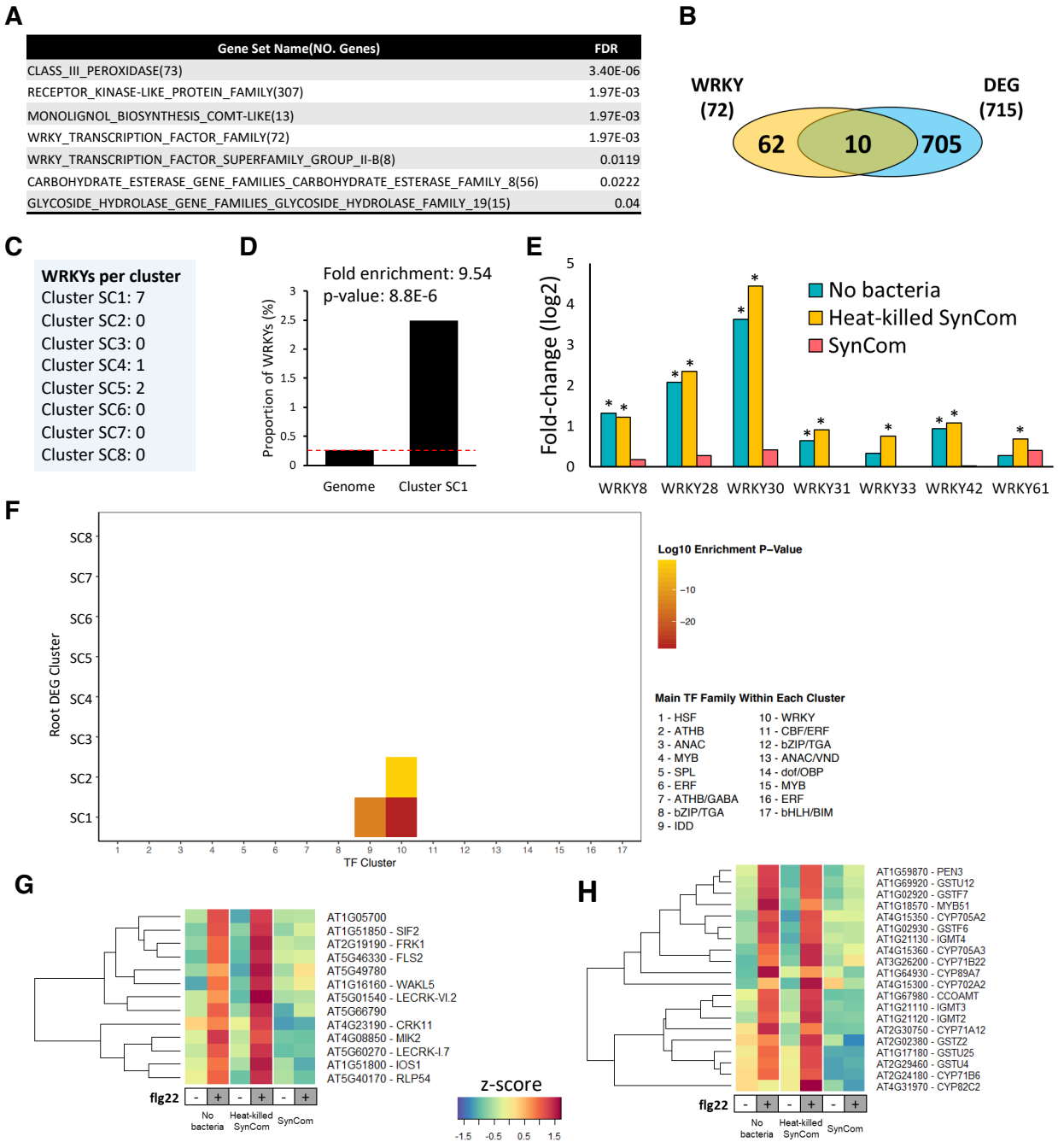


Figure 4.7. SynCom35 suppresses key components of the plant immune system. (A) Gene families enriched among the group of flg22-responsive genes that were suppressed by SynCom35 (cluster SC1; Fig. 4.1F). Note the prevalence of defense-related families. (B) A total of 72 genes encoding WRKY transcription factors were annotated in the Arabidopsis Col-0 ecotype at The Arabidopsis Information Resource (TAIR). Among these, 10 were differentially expressed in the roots of plants treated with flg22 in our study. (C) Distribution of WRKYs per cluster (Fig. 4.1F). Eight of the ten flg22-responsive WRKY genes were up-regulated by flg22 (clusters SC1 and cluster SC4), and two were down-regulated (cluster SC5). Note that seven of the eight WRKYs that are activated by flg22 are included in cluster SC1 (i.e., they are suppressed by SynCom35).

(D) Enrichment analysis showing that cluster SC1 contains approximately 10 times more WRKY genes than expected by chance (hypergeometric test; p-value = 8.8E-6). (E) Fold-changes (flg22 vs mock) of the seven WRKY genes included in cluster SC1. Note that these genes are not activated by flg22 when living SynCom35 is also present. Asterisks indicate genes that are differentially expressed based on the edgeR analysis. (F) Genes containing WRKY-binding motifs in their promoters (i.e., putative WRKY targets) are highly enriched in cluster SC1 and, to a lesser extent, in cluster SC2. (G) Expression profile of receptor kinase-like genes found in cluster SC1 (Fig. 4.1F). FRK1 is a widely used as a marker of MTI and FLS2 is the flg22 receptor. (H) Expression profile of genes involved in the synthesis of secondary metabolites found in cluster SC1 (Fig. 4.1F).

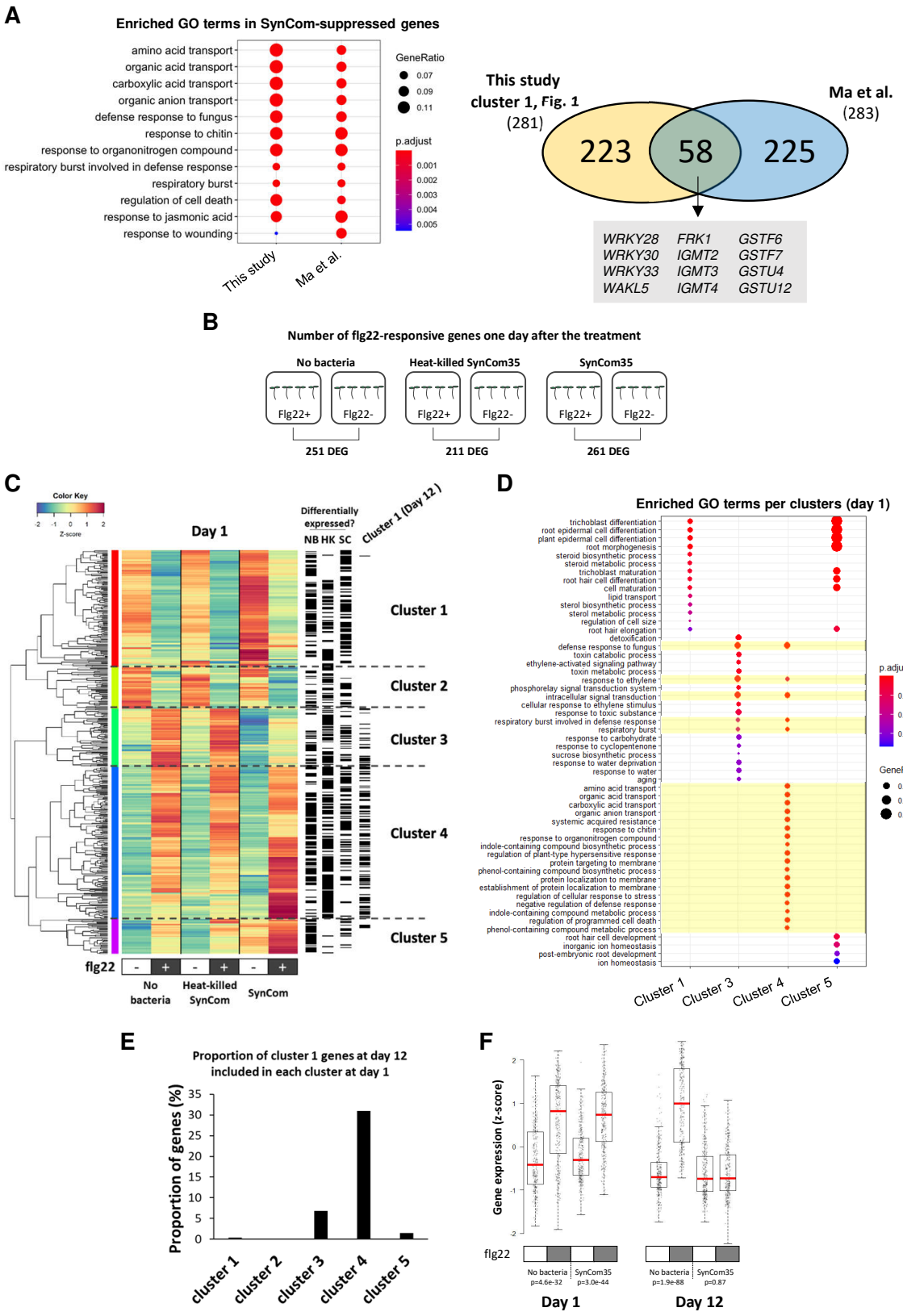


Figure 4.8. SynCom35 actively suppresses an ongoing immune response that is also observed in other plant-microbiota interactions. (A) Suppression of defense-related genes is

also observed with an independently derived SynCom (Ma et al., 2020). The left side shows that the genes suppressed by the SynCom reported by Ma et al. are enriched in the same biological processes identified in the genes suppressed by SynCom35 in this study (cluster SC1, Fig. 4.1F). p-values were adjusted with the FDR method. The right side shows the overlap between the two sets of genes, highlighting 58 genes that are commonly suppressed by the SynComs in these independent studies (p=2.53E-58; hypergeometric test). Representative shared genes are shown in the grey box. The complete list of shared genes is shown in SI Appendix, Dataset S3. (B) Number of differentially expressed genes one day after the treatment of roots with flg22 (or mock) in the absence of bacteria, with heat-killed SynCom35 or with SynCom35 alive. (C) Hierarchical clustering of the 449 genes that responded to flg22 in at least one of the experimental conditions one day after treatment. In contrast to the results observed 12 days after treatment (Fig. 4.1F), SynCom35 does not cause a major interference in the root response to flg22 at day 1, indicating that SynCom35 interferes with an ongoing immune response. Indeed, flg22-responsive genes that are suppressed by SynCom35 at day 12 (Cluster 1, Fig. 4.1F) are activated at day 1 (black marks in the "Cluster 1" lane on the right of the heatmap). Genes that respond to flg22 significantly in each of the treatments at day 1 according to edgeR are indicated with black marks on the right in the "Differentially expressed" lanes (NB: no bacteria; HK: heat-killed SynCom35; SC: SynCom35). (D) Gene ontology enrichment analyses showing enriched biological processes in each of the five clusters. Cluster 4, which contains defense-related genes, is highlighted in yellow. No biological processes were enriched in Cluster 2. p-values were adjusted with the FDR method. (E) Proportion of SynCom35- suppressed genes at day 12 (Cluster 1; Fig. 4.1F) that is present in each cluster at day 1. Note that over 30% of the genes suppressed by SynCom35 at day 12 are included in Cluster 4 at day 1. (F) Expression profile of the 281 flg22-responsive genes that are suppressed by SynCom35 at day 12 (Cluster SC1; Fig. 4.1F). Note that the expression of these genes is increased by flg22 in the absence of bacteria at both time points (1 and 12 days after treatment). However, their expression is reduced in the presence of SynCom35 only 12 days after treatment. The p-values on the bottom refer to the comparison between groups (flg22 vs mock) for each condition using t-tests.

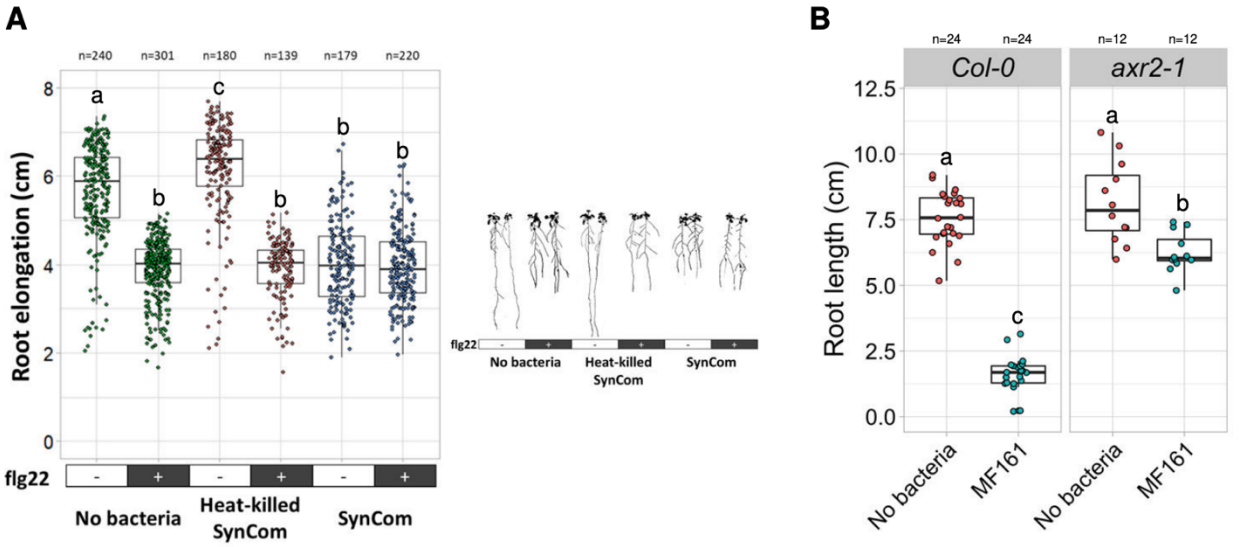


Figure 4.9. SynCom35 effects on the Arabidopsis root morphology. (A) Root elongation measured in Arabidopsis seedlings grown in the absence of bacteria, with heat-killed SynCom35 and with SynCom35. For each of these three conditions, plants were exposed or not to 1 µM flg22, resulting in six treatments. Root growth inhibition (RGI) was induced by flg22 in plants grown in the absence of bacteria or with heat-killed SynCom35. Seedlings grown with SynCom35 displayed shorter roots even in the absence of flg22. Multiple comparisons were performed with ANOVA followed by a Tukey test (a = 0.05). Representative images of seedlings from each treatment are shown on the right. (B) The auxin-producing strain Arthrobacter sp. MF161 included in SynCom35 induces the RGI phenotype when in mono-association with Arabidopsis. This phenotype is significantly reduced in the auxin-resistant mutant line axr2-1 suggesting that it is mediated by an auxin response. Multiple comparisons were performed with ANOVA followed by a Tukey test (a = 0.05).

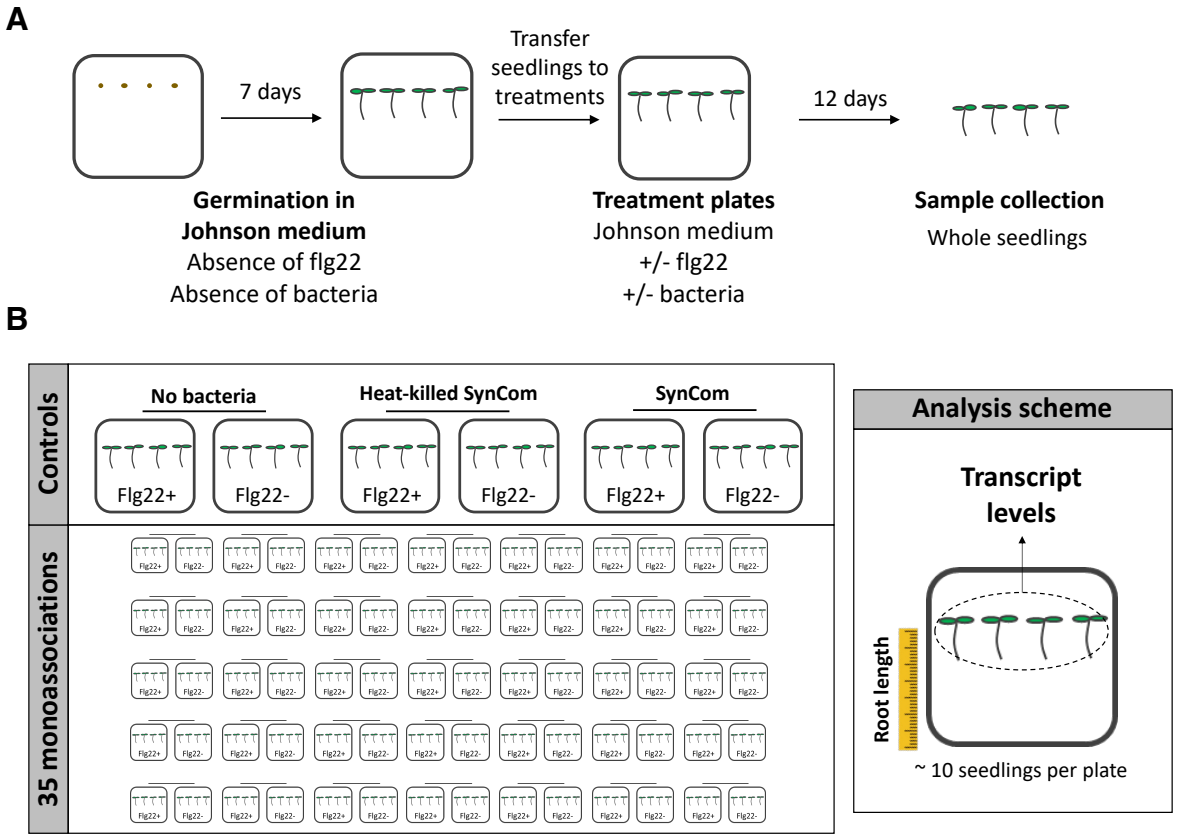


Figure 4.10. Design of the mono-association experiment. (A) Representation of the experimental system used to evaluate the effect of each of the 35 members of SynCom35 on Arabidopsis seedlings. Seeds of wild-type Arabidopsis (Col-0 ecotype) were germinated in Johnson medium under gnotobiotic conditions in the absence of flg22. After seven days, the resulting seedlings were transferred to plates (10 per plate) containing Johnson medium embedded with a single bacterial strain (105 cfu/mL) in the presence or not of 1 µM flg22. Whole seedlings were harvested for RNA extraction twelve days later. Plants grown in the absence of bacteria and with heat-killed SynCom35 and SynCom35 were used as controls. (B) Representation of the conditions included in the experiment. A total of 76 different conditions were evaluated (35 strains and 3 controls x 2 treatments) in three independent experiments, each containing 3 biological replicates (total of 9 biological replicates per condition). The root length and the plant transcriptional program (RNA-seq) was assessed in each sample.

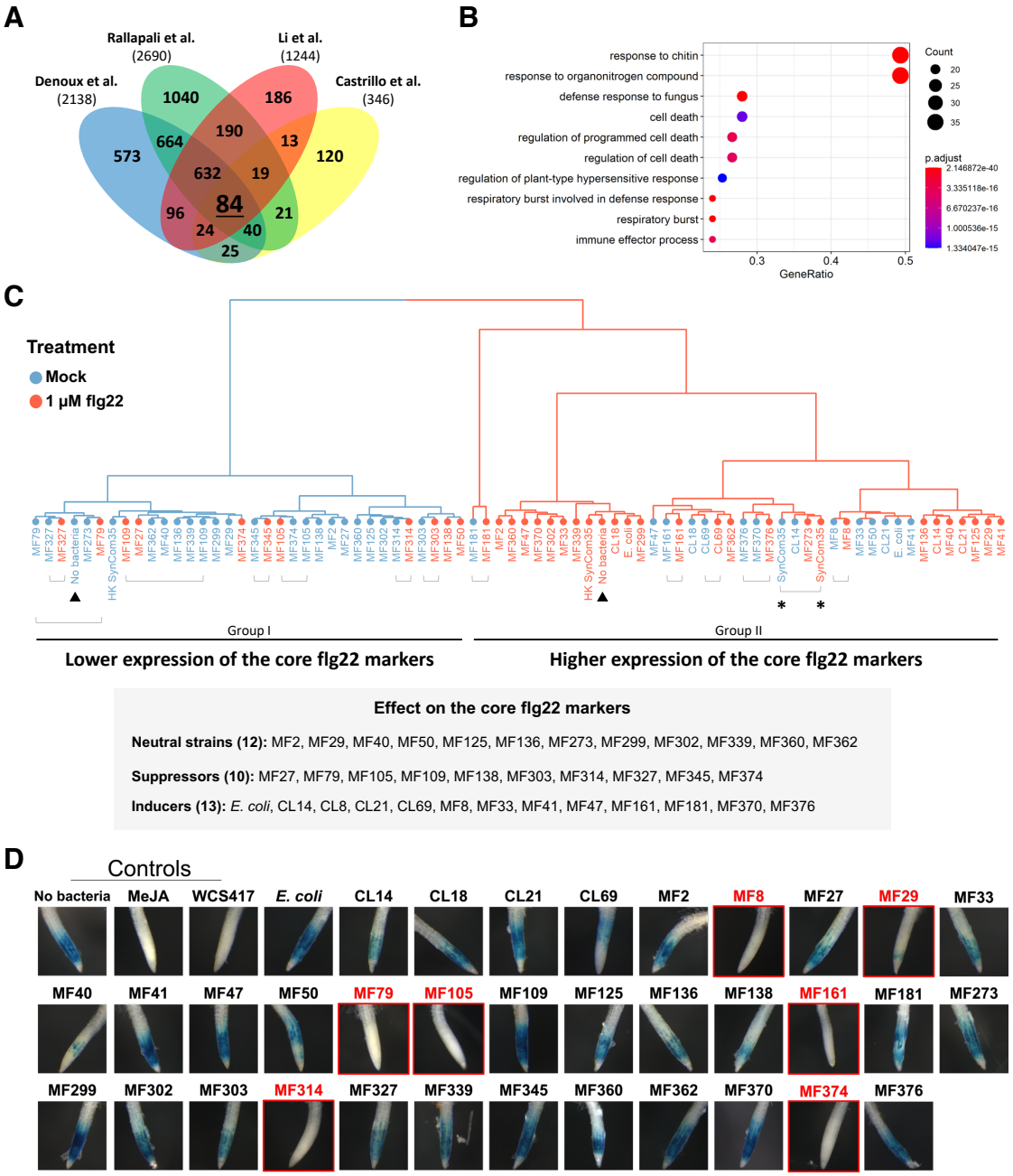


Figure 4.11. Members of SynCom35 interfere with the expression of a robust set of 84 core flg22-responsive genes in a chronic treatment and also suppress the plant acute response to the elicitor. (A) Venn diagram with the overlap among genes found up-regulated by flg22 in four different gene expression experiments (1, 38–40). The intersection (84 genes; SI Appendix, Dataset 5) was used as a robust marker set of the flg22 response. (B) A Gene Ontology enrichment analysis showed that this core set of genes is highly enriched in immunity-related biological processes. p-values were adjusted with the FDR method. (C) Plants grown in mono-association with SynCom35 members were clustered based on the expression of the 84 markers genes of the flg22 response (core flg22 markers). Plants grown with SynCom35 (asterisks), heat killed SynCom35 and in the absence of any bacteria ("no bacteria"; triangles) were also included in the experiment. For each condition, plants were exposed to 1 µM flg22 (red labels) or a mock

treatment (blue labels). Note that samples are organized in two main clusters: the cluster on the left includes most of the mock treated plants (blue) and displays lower expression of the core flg22 marker set. The cluster on the right includes most of the flg22-treated plants (red) and displays higher expression of the same marker set. In the presence of some strains, plants treated with flg22 (red) displayed low expression of the flg22 response marker set and, thus, grouped with mock-treated samples on the left. These were considered suppressor strains. In other cases, bacteria triggered higher expression of the flg22 marker set even in the absence of exogenous MAMP (blue), grouping with flg22-treated plants on the right. These were considered inducer strains. Lines under the labels connect plants treated with a specific strain for which addition of flg22 had no major effect on the expression of the 84 marker genes (i.e., mock- and flg22-treatments grouped together). The chart below lists strains with neutral, suppressive or inducive effect on the flg22 markers. (D) At least seven members of SynCom35 (labeled in red) suppress the acute flg22 response in Arabidopsis roots. The reporter line carrying the pCY71A12::GUS construct (25) was grown for seven days in gnotobiotic conditions and then inoculated with each single strain individually at a final OD of 0.002. After 14h of co-cultivation, plants were treated with 100 nM flg22 for 5h. A blue signal in the elongation zone indicates that the pCYP71A12 was activated by flg22. MeJA and Pseudomonas simiae WCS417 were previously shown to suppress MTI in the roots and were used as controls in the experiment (25). The experiment was performed at least two independent times for each strain. Each time included 2 repetitions (wells) consisting of approximately 10 seedlings each.

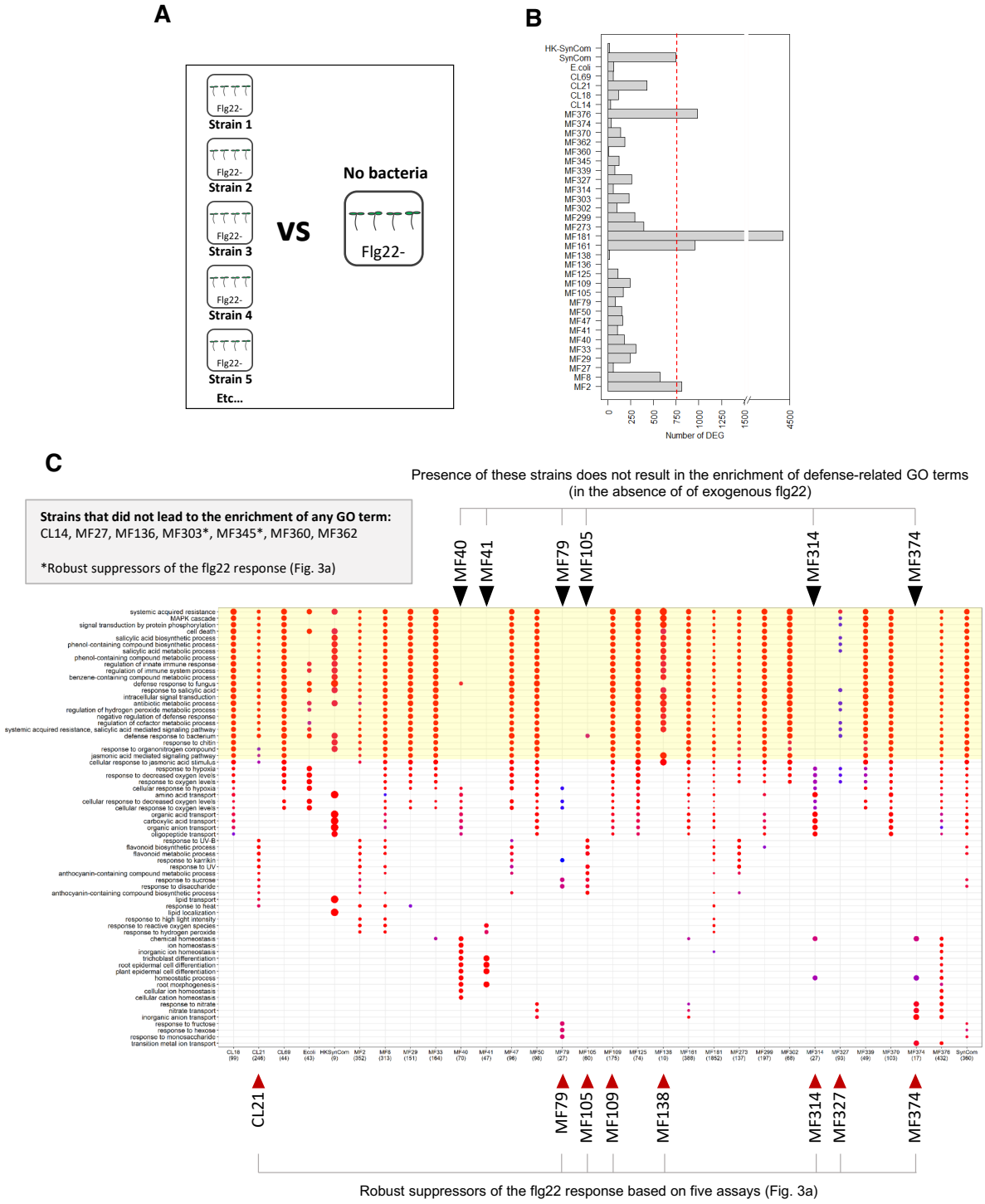


Figure 4.12. Most, but not all, members of SynCom35 trigger the activation of plant defense genes to some extent when in mono-association. (A) Representation of the comparisons performed to evaluate the effect that each member of SynCom35 has on the plant. The transcriptome of plants grown with individual strains in the absence of flg22 was compared to the control condition (no addition of bacteria or flg22). (B) Number of differentially expressed genes identified in plants grown with each SynCom35 member. The red line depicts the number of differentially expressed genes identified in plants grown in the presence of SynCom35. (C) Gene

Ontology (GO) enrichment analysis showing biological processes enriched in the sets of genes upregulated by each strain. The area shaded in yellow corresponds to GO terms related to immune responses. Note that most strains trigger a transcriptional program that is enriched in defense-related genes. The six strains indicated with arrows on the top (MF40, MF41, MF79, MF105, MF314 and MF374) do not lead to the activation of the plant immune response but trigger the activation of other biological processes. Seven other strains (MF27, MF136, MF303, MF345, MF360, MF362 and CL14) did not result in the enrichment of GO processes and are not shown in the figure. Strains defined as robust suppressors (Fig. 4.3A) are labeled on the bottom. p-values were adjusted with the FDR method.

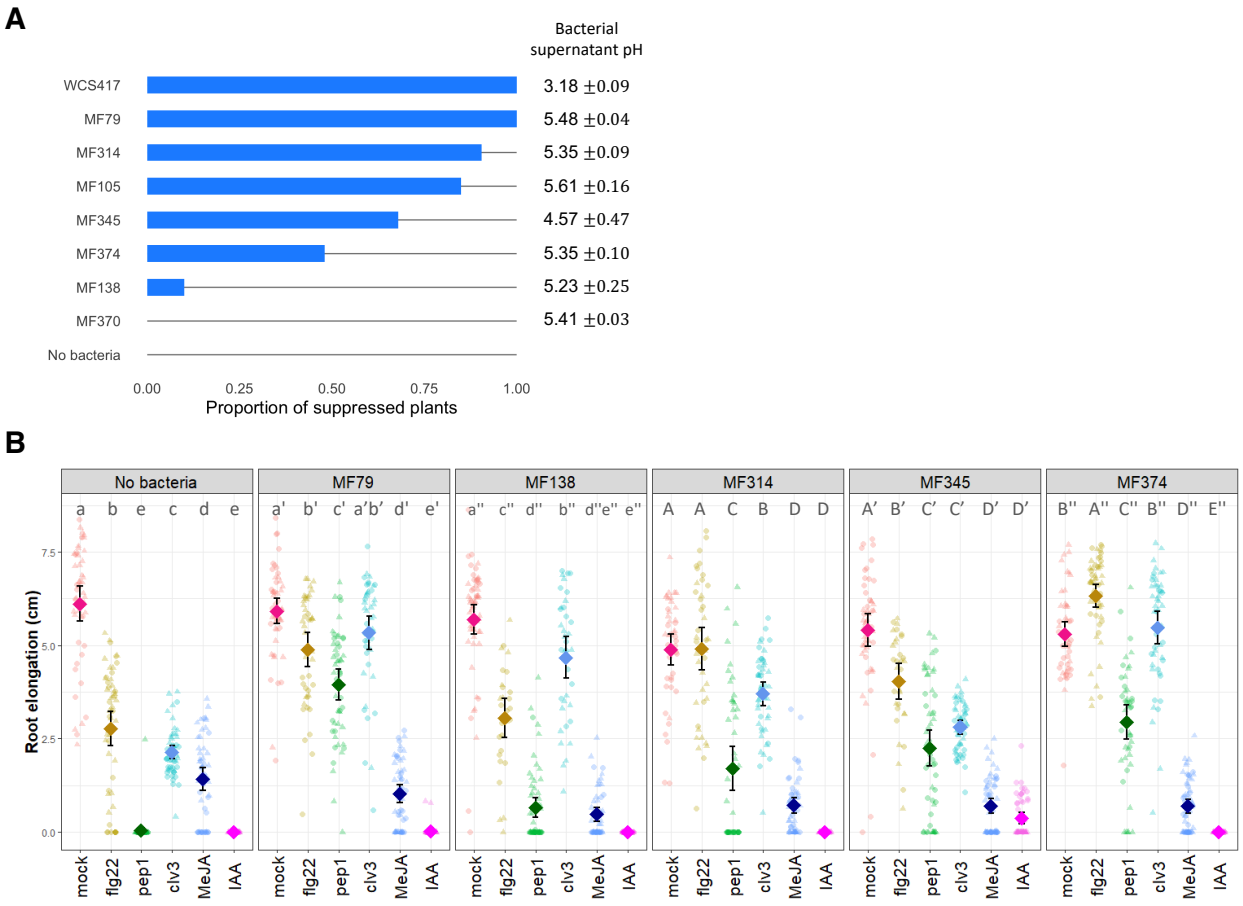


Figure 4.13. Suppression of the plant immune system by commensal bacteria likely occurs through multiple different mechanisms. (A) Six robust suppressor strains were evaluated for their ability to prevent the root response to flg22 through acidification of the extracellular medium. The assay was performed as described by Yu et al. (2019) using cell-free bacterial supernatants and the Arabidopsis pCYP71A12::GUS reporter line. Pseudomonas simiae WCS417, which suppresses the flg22 response through medium acidification, was used as a positive control. The non-suppressor strain Ochrobactrum sp. MF370 was used as a negative control. The bars show the proportion of plants that did not respond to flg22 in each treatment (i.e., proportion of suppressed plants). A total of 20-25 seedlings were evaluated in the treatment of each suppressor strain. Note that the cell-free supernatant of most strains was sufficient to suppress the root response to flg22. Consistent with previous observations (Yu et al., 2019), P. simiae WCS417 lowered the

extracellular pH to ~3. However, the supernatant of our suppressor strains remained at pH of 5-6, which is sufficient for a normal response to flg22 (Yu et al., 2019). This indicates that these strains do not employ the acidification of the extracellular medium as a strategy to suppress MTI. (B) Evaluation of the ability of suppressor strains to interfere with the root growth inhibition (RGI) phenotype induced by various molecules. The peptides flg22 (1 μM), pep1 (1 μM) and clv3 (1 μM) as well as the hormones methyl jasmonate (MeJA; 10 μM) and indole-3-acetic acid (IAA; 1 μM) were used. Seven-day-old Col-0 seedlings were transferred to Johnson medium embedded with bacteria (OD600nm=0.0001) and the molecules of interest and grown at 9h light/15h dark (21oC light/18oC dark). Root elongation was Note that all molecules induce RGI in plants grown in the absence of bacteria (control treatment). measured after 12 days. Most strains suppress the RGI phenotype caused by peptides but not by the two hormones. Diamonds represent means with two times standard error. Multiple comparisons within each 'bacteria' condition were performed with ANOVA followed by a Tukey test (letters on the top).

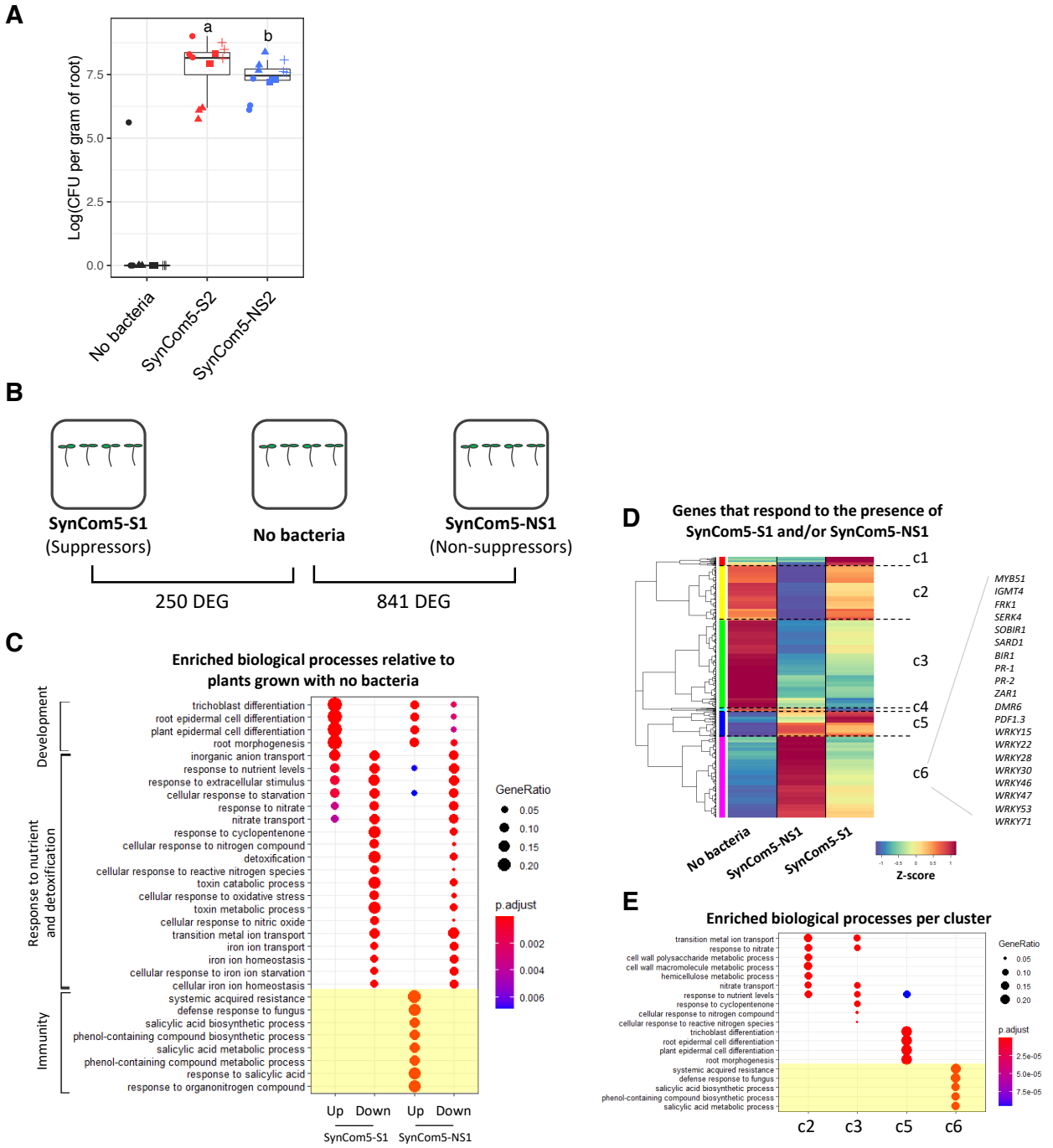
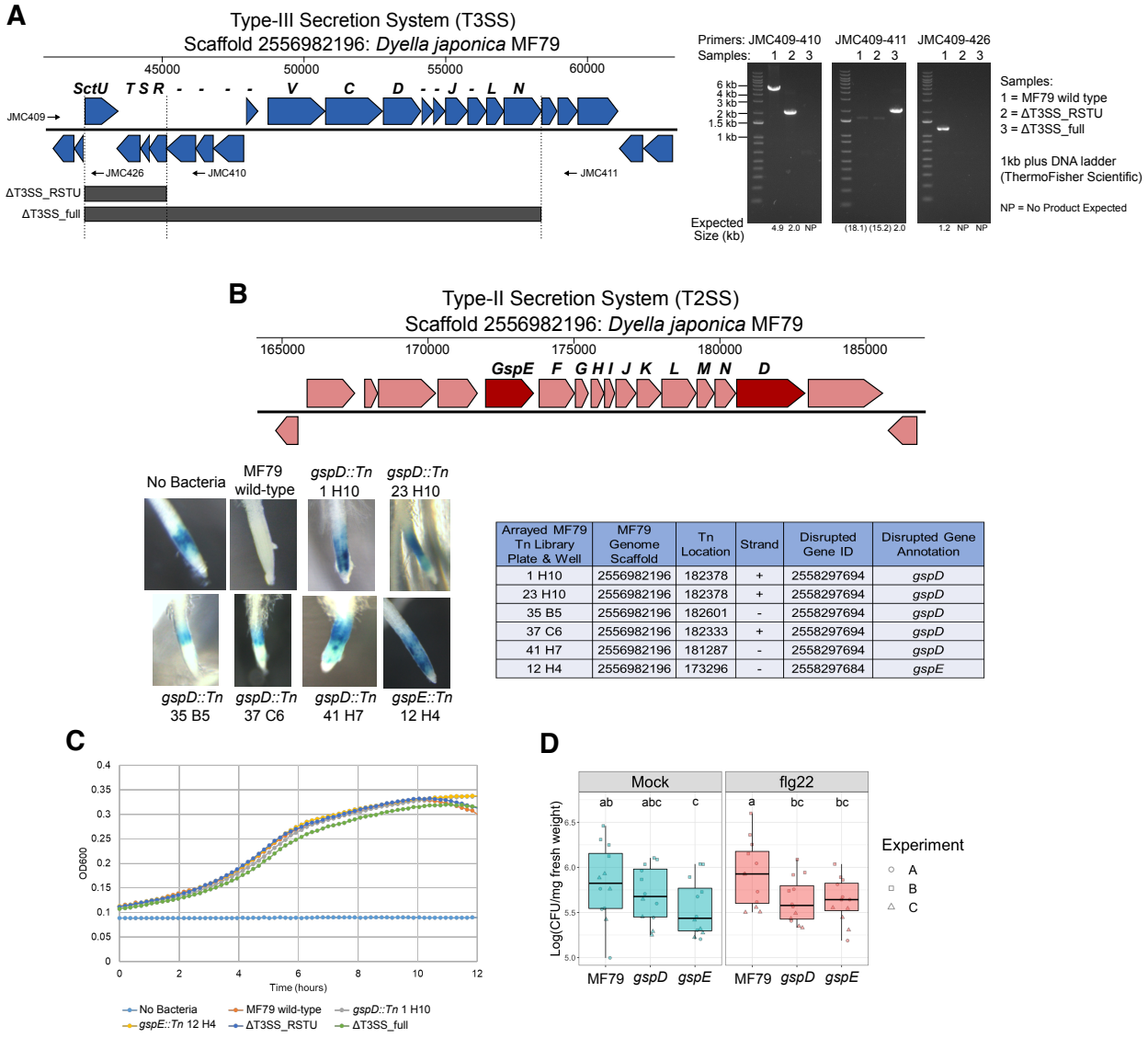


Figure 4.14. A community of non-suppressor bacteria triggers defense responses in Arabidopsis. (A) Evaluation of bacteria colonization of Arabidopsis roots. A SynCom made of five suppressors (SynCom5-S2) grows to higher levels in the roots than a SynCom made of five non-suppressors (SynCom5-NS2). The experiment was performed four independent times, each with three biological replicates per condition (n=12). Different symbols represent different experimental repetitions. This experiment complements the results presented in Fig. 4.3C and was performed in a different experimental system (see methods) using different SynComs. (B) Transcriptional analysis of plants grown in the absence of bacteria or with communities of suppressors (SynCom5-S1) and non-suppressor (SynCom5-NS1) strains. The number of

differentially expressed genes (DEG) is shown under each comparison. No exogenous flg22 was added. (C) Gene ontology enrichment analysis showing biological processes enriched in the sets of genes up- or down-regulated in plants grown with SynCom5-S1 or SynCom5-NS1 relative to axenic conditions. Note the absence of immunity related terms in plants colonized with SynCom5-S1 (shaded in yellow). (D) Hierarchical clustering of 925 genes that responded to SynCom5-S1, SynCom5-NS1 or both. Six clusters were defined. Example of defense-related genes are shown on the right. (E) Gene ontology enrichment analysis of each cluster defined in panel (D). Note that immunity-related terms are enriched in cluster c6, which includes genes more expressed in the presence of SynCom5-NS1. Clusters c1 and c4 do not have enriched terms. SI Appendix, Dataset 6 presents metadata information of the experiment, the lists of differentially expressed genes and the complete results of the enrichment analyses.



Arrayed MF79 Tn Library Plate & Well	MF79 Genome Scaffold	Tn Location	Strand	Disrupted Gene ID	Disrupted Gene Annotation
1 H10	2556982196	182378	+	2558297694	*gspD*
23 H10	2556982196	182378	+	2558297694	*gspD*
35 B5	2556982196	182601	-	2558297694	*gspD*
37 C6	2556982196	182333	+	2558297694	*gspD*
41 H7	2556982196	181287	-	2558297694	*gspD*
12 H4	2556982196	173296	-	2558297684	*gspE*

Figure 4.15. The T2SS and not the T3SS of *Dyella japonica* MF79 is required for suppression of the flg22 response. (A) Gene map for the *Dyella japonica* MF79 type-III secretion system (T3SS, Sct genes). DNA coordinates on Scaffold 2556982196 are identified above the map. Genes deleted in the MF79 ΔT3SS_RSTU and MF79 ΔT3SS_full mutants are marked below the map.

DNA gels confirming gene knockouts and strain purity in these mutants are shown to the right. Expected product sizes are shown below the gels. The location of primers used to screen these mutants are marked in the T3SS map. Primers sequences are available in SI Appendix, Dataset 7. (B) Gene map for the Dyella japonica MF79 type-II secretion system (T2SS, Gsp genes). Six strains with loss of flg22-response suppression were identified from the arrayed MF79 transposon insertion library. Transposon insertion locations were identified by arbitrary PCR. At left, GUS expression assay demonstrating loss of flg22-response suppression by the 6 strains (5 insertions in GspD, 1 insertion in GspE). Genes disrupted in the MF79 gspE::Tn and gspD::Tn strains are highlighted in dark red in the map of the T2SS. (C) Growth of the MF79 wild-type, T3SS deletion, and T2SS transposon insertion mutant strains in Minimal Medium A with 10 mM sucrose and 0.1% tryptone. (D) The ability to suppress the plant immune system correlates with enhanced colonization of Arabidopsis roots. Although gspD and gspE mutants did not display impaired growth in vitro, they are poorer colonizers of Arabidopsis roots than the wild-type strain. Plants were cultivated with each strain in mono-colonization for 5 days in the presence (red) or not (blue) of 1 μM flg22.

Section 4.5: Methods

Preparation of bacterial synthetic communities (SynComs). We employed SynComs to evaluate the effect of the root microbiota on plant immune responses. A 35-member SynCom (SynCom35) was assembled using a set of genome-sequenced strains that represent the typical taxonomic diversity of plant-associated bacteria (38). This SynCom included 32 strains that were isolated from Brassicaceae roots (mainly Arabidopsis), two strains that were isolated from unplanted soil (Mason Farm - North Carolina, USA; +35° 53′ 30.40″, −79° 1′ 5.37″) and Escherichia coli DH5α as a control. Smaller SynComs were assembled with subsets of strains from this 35- member SynCom (i.e., SynCom5-S: five suppressors; and SynCom5-NS: five non-suppressors). A detailed description of the strains that comprise the SynComs used in this study is provided in SI Appendix, Dataset 1.

Assembly of SynComs followed the procedures described by Castrillo et al. (38). Briefly, a single colony of each strain was inoculated into 4 mL 2xYT medium (16 g/L tryptone, 10 g/L yeast extract and 5 g/L NaCl) in a test tube and kept at 28oC under agitation at 250 rpm. After one to two days, bacterial cultures were washed twice with 10 mM MgCl2 (by centrifugation and

resuspension) and OD600nm was determined by spectrophotometry. Cultured bacterial strains were then added to 500 mL of medium to a final concentration of 105 cfu/mL each (assuming that one OD600nm unit equals to 109 cfu/mL). In some experimental conditions, the bacterial elicitor flagellin 22 (flg22) was added along with bacteria at a final concentration of 1 μM (described below).

Preparation of plants for bacteria and flg22 treatments. Arabidopsis thaliana seeds (ecotype Columbia; Col-0) were surface-sterilized with a cleaning solution (70% bleach and 0.2% Tween20) for 8 min and rinsed three times with sterile distilled water to eliminate any seed-borne microbes on the seed surface. After stratification at 4°C for 2 days in the dark, seeds were transferred to square plates containing Johnson medium (5 g/L sucrose; 85 mg/L KH2PO4; 0.6 g/L KNO3; 0.9 g/L Ca(NO3)2·4H2O; 0.2 g/L MgSO4·7H2O; 3.8 mg/L KCl; 1.5 mg/L H3BO3; 0.8 mg/L MnSO4·H2O; 0.6 mg/L ZnSO4·7H2O; 0.1 mg/L CuSO4·5H2O; 16.1 μg/L H2MoO4; 1.1 mg/L FeSO4·7H2O; 0.1 g/L MyoInositol; 0.5 g/L MES; pH 5.6–5.7) solidified with 1% bacto-agar. Plates were kept in the vertical position for seven days in a growth chamber under a 16-hour light/8-hour dark regime at 21oC day/18oC night. The resulting seedlings were then ready for experiments.

Treatment of plants with flg22 in the presence of SynComs. We evaluated the effect of SynCom35 on the plant response to flg22 (Fig. 4.1) by cultivating plants in Johnson medium with or without 1 μM of flg22 (synthesized by PhytoTech Labs) in the presence or not of the SynCom (inoculum: 105 cfu/mL of each strain; prepared as described above). A control condition in which plants were exposed to the heat-killed SynCom was included in the experiment by heating the bacteria at 100°C for 2h in an oven immediately before inoculating

them into the medium at the same concentration as the SynCom alive. The complete design of this experiment as well as the parameters measured are shown in SI Appendix, Fig. S1.

Seven-day-old seedlings were obtained by germinating the seeds in Johnson medium as described above and then transferred to the six different experimental conditions (SI Appendix, Fig. S1). Plates were kept in a growth chamber under a 9-hour light/15-hour dark regime at 21oC day/18oC night. The entire root system of the seedlings was harvested for RNA extraction at two time-points: 1 and 12 days after the transfer to the treatment plates. Tubes were stored at −80oC until processing. The experiment included 9 biological replicates per condition, which were sub-divided into 3 batches (each containing 3 replicates per condition) that were performed independently (i.e., different days, medium, seedlings and bacterial inoculum). Each replicate corresponds to the root systems of ten seedlings grown in a single plate.

Treatment of plants with flg22 in the presence of individual strains. We evaluated the effect of each member of SynCom35 on the plant response to 1 μM flg22 in the context of monoassociations (Fig. 4.2 and SI Appendix, Fig. 4.10). For this, each strain was cultured and added individually to 500 mL of medium as described above at a concentration of 105 cfu/mL. Treatment of seedlings with flg22 in mono-association was performed as described for SynCom35 (see above). Seven-day-old seedlings were transferred to the different conditions of the experiment and harvested after 12 days for RNA extraction. Tubes were stored at −80oC until processing. The experiment included 9 biological replicates per condition, which were sub-divided into 3 batches (each containing 3 replicates per condition) that were performed independently (i.e., different days, medium, seedlings and bacterial inoculum). Each replicate corresponds to ten seedlings grown in a single plate. SI Appendix, Fig. 4.10 summarizes the design of this experiment.

Root length measurements. Before harvesting samples for RNA extraction at day 12 (end of the experiments), plates were scanned, and the root length of each seedling (10 per plate) was determined using ImageJ (39). Results from the SynCom35 experiment (Fig. 4.1 and SI Appendix, Fig. 4.9) were visualized with a boxplot made in R using the ggplot2 package (40). Due to the large number of treatments (i.e., strains) in the mono-association experiment (Fig. 4.2), results were visualized with a heatmap. For this, the raw root length data was fit with a linear model using the lm function in R. We identified a significant batch effect using the model:

$$Root\_length \sim Bacteria * flg22 + Batch$$

"Root_length" is the root length from each replicate, "Bacteria" is the strain inoculated into the plant growth medium, "flg22" is the presence of 1 μM flg22 or not, and "Batch" is the experimental effect (i.e., each of three independent batches of the experiment). To correct the root lengths for the batch effect, we used the beta effects of bacteria, flg22, and bacteria:flg22 from the linear model to calculate fitted root length measurements for each bacterial strain, one with flg22 and one without. These fitted measurements were then clustered using a Euclidean distance matrix and the ward.d2 clustering method implemented in the hclust function in R. The data was plotted with the R package ComplexHeatmap (41). The heatmap was annotated to indicate the Class of each strain as well as the presence or not of a type-3 secretion system (T3SS), which was predicted using macsyfinder and hidden Markov Models described by Abby et al. (42). The code used for this analysis can be found at GitHub (https://github.com/ncolaian/MAE).

RNA extraction and RNA-seq library preparation. Total RNA was extracted following the procedures described by Logemann et al. (43). Frozen plant tissue was pulverized inside a 2 mL tube containing three 4 mm glass beads using a Qiagen TissueLyser II instrument.

Samples were homogenized in 400 μL of Z6-buffer (8 M guanidinium- HCl, 20 mM MES, 20 mM EDTA, pH 7.0). Following the addition of 400 μL phenol:chloroform:isoamyl alcohol (25:24:1), samples were vortexed and centrifuged (20,000 g, 10 min) for phase separation. The aqueous phase was transferred to a new 1.5 mL tube and 0.05 volumes of 1 N acetic acid and 0.7 volumes 96% ethanol were added. The RNA was precipitated at −20 °C overnight. Following centrifugation (20,000 g, 10 min, 4 °C) the pellet was washed with 200 μL sodium acetate (pH 5.2) and then 70% ethanol. The RNA was dried and dissolved in 30 μL of ultrapure water and stored at −80 °C until use.

Illumina mRNA-seq libraries were prepared from 1,000 ng RNA using the protocol described by Castrillo et al. (38). Briefly, mRNA was purified from total RNA using Sera-mag oligo(dT) magnetic beads (GE Healthcare Life Sciences) and then fragmented in the presence of divalent cations (Mg2+) at 94oC for 6 min. The resulting fragmented mRNA was used for first-strand cDNA synthesis using random hexamers (ThermoFisher Scientific) and the EnzScript reverse transcriptase (Enzymatics), followed by second strand cDNA synthesis using DNA polymerase I (Enzymatics) and RNaseH (Enzymatics). Double-stranded cDNA was end-repaired using T4 DNA polymerase (Enzymatics), T4 poly-nucleotide kinase (Enzymatics) and Klenow polymerase (Enzymatics). The DNA fragments were then adenylated using Klenow exo-polymerase (Enzymatics) to allow the ligation of Illumina Truseq HT adapters (D501–D508 and D701–D712). Following library preparation, quality control and quantification were performed using a 2100 Bioanalyzer instrument (Agilent) and the Quant-iT PicoGreen dsDNA Reagent (ThermoFisher Scientific), respectively. Libraries were sequenced on the Illumina HiSeq4000 instrument to generate 50-bp single-end reads. A total of 100 and 666 libraries were produced for

the SynCom35 (Fig. 4.1) and the mono-association (Fig. 4.2) experiments, respectively. SI Appendix, Datasets 3 and 4 present details about each sample included in these experiments.

Processing of RNA-seq reads. Initial quality assessment of raw sequences was done with FastQC version 0.11.7 (Babraham Bioinformatics, Cambridge, UK). Trimmomatic version 0.36 (44) was used to remove adaptor-containing and low-quality sequences with parameters set at ILLUMINACLIP:TruSeq3-SE.fa:2:30:10, SLIDINGWINDOW:4:5, LEADING:5, TRAILING:5, MINLEN:50. The resulting high-quality reads were then aligned against the TAIR10 Arabidopsis reference genome using HiSat2 version 2.1.0 (45) using default parameters. The featureCounts function from the Sub-read package version 1.6.3 (46) was used to count reads that mapped to each one of the 27,206 nuclear protein-coding genes. Evaluation of the results of each step of the analysis was done with MultiQC version 1.7 (47). Raw RNA-seq reads and the count matrices generated in this study are available at the NCBI Gene Expression Omnibus under the accession number GSE156426.

Evaluation of the transcriptional response to flg22 in plants growing in the presence of SynComs. Identification of differentially expressed genes was performed using the generalized linear model (glm) approach (48) implemented in the R package edgeR (49). In the SynCom35 experiment (Fig. 4.1), weakly expressed genes were filtered out by removing those genes that did not achieve a minimum expression level of 1 count per million in at least five libraries. Normalization was performed using the trimmed mean of M-values method (TMM; function calcNormFactors in edgeR) (50). The glmFit function was used to fit the counts to a negative binomial generalized linear model with a log link function. In this experiment, our goal was to evaluate the root transcriptional response to flg22 in plants grown in the presence of the SynCom35 (alive or dead) or in the absence of bacteria (Fig. 4.1B). For this, we employed a one-

way-layout model to compare each treatment ('no bacteria.flg22', 'SynCom35.flg22' or 'heat-killed SynCom35.flg22') to its control condition ('no bacteria.mock, 'SynCom35.mock' or 'heat-killed SynCom35.mock') at each time-point (T1h and T12h). The model included the covariate 'Experiment' to control for batch effects associated with the three independent repetitions of the experiment. The Benjamini–Hochberg method (false discovery rate; FDR) was used for the correction of multiple comparisons (51). Genes with an FDR corrected p-value below or equal to 0.01 and a fold-change of at least 1.5x were considered differentially expressed. We compared the sets of differentially expressed genes in each condition ('no bacteria', 'SynCom' and 'Heat-killed SynCom') using a Venn diagram (52) (Fig. 4.1C). Pearson correlations of flg22 responses (Fig. 4.1D) were estimated by comparing the fold-changes of all 715 genes that responded to flg22 in any of our conditions using the R package corrplot (53). A Constrained Analysis of Principal Coordinates (CAP) was performed to allow the visualization of the relative effects of our experimental conditions on the plant transcriptome (Fig. 4.1E). For this, we selected the 500 genes with the highest standard deviation across all samples from the matrix of normalized gene counts generated by edgeR and used the capscale function from the vegan package (54) to compute the CAP. We constrained the ordination on the presence of the SynCom and the addition of flg22, while conditioning for the batch effect. Hierarchical clustering (Fig. 4.1F and SI Appendix, Fig. 4.8C) was performed with the 'heatmap.2′ function from the R package gplots (55) using those genes that responded to flg22 in at least one condition of the experiment (i.e, 'no bacteria', 'SynCom35' or 'Heat-killed SynCom35') within each time-point. Genes were clustered on the basis of the Euclidean distance with the complete-linkage method. The cutree function was used to define eight clusters based on dendrogram distances. To generate a representative view of each cluster (Fig. 4.1G), RPKM expression values were normalized by z-

score transformation and presented in a boxplot. Gene Ontology (GO) enrichment analyses (Fig. 4.1H and SI Appendix, Figs 4.6 and 4.8) were performed with the compareCluster function of the R package clusterProfiler (56) and with the PlantGSEA platform (57).

Evaluation of the transcriptional response to flg22 in plants growing in the presence of individual strains. In the mono-associations experiment (Fig. 4.2), weakly expressed genes that did not have greater than 1 count per million in at least five libraries were removed from subsequent analyses. We then normalized the gene count matrix using the weighted trimmed mean of M-values (TMM) implemented in the calcNormFactors function from edgeR. We were interested in identifying (38) the plant genes that respond to bacteria in mono-association (bacteria_mock vs axenic_mock); and (39) plant genes that respond to flg22 in the presence of each strain in mono-association (bacteria_flg22 vs bacteria_mock) or in the 'no bacteria' control (axenic_flg22 vs axenic_mock). To perform these comparisons, we first subset the raw gene count table to only include the conditions being tested and created a DGEList object with this data using edgeR. We normalized the counts based on normalization factors calculated from the full gene count table. We then estimated the gene-wise dispersion parameters within each of the smaller datasets using the estimateGLMCommonDisp, estimateGLMTrendedDisp and estimateGLMTagwiseDisp functions from edgeR. Our design matrices were as follows for each of the contrasts listed above:

(1) expression ~ bact*experiment

(2) expression ~ treatment+experiment

"bact" is the bacterial strain, "experiment" is the batch effect (i.e., each of the three independent repetitions of the experiment), and "treatment" is the flg22 status. We then fit a gene-wise negative binomial generalized linear model with our calculated dispersion parameters using the

glmFit function. After fitting the models, we conducted gene-wise log ratio tests contrasting our factor of interest, (1) bacterial strain and (2) flg22 added, to the controls: (1) 'no bacteria' and (2) 'no flg22 added' using the glmLRT function. We used the decideTestsDGE function in edgeR to calculate FDR corrected p-values and report differentially expressed genes that had a log fold-change of at least 1.5 and an FDR corrected p-value of less than 0.01.

Principal Component Analysis (PCA) was performed with the set of 428 flg22-responsive genes defined in the control condition of the experiment (i.e., plants growing in the absence of bacteria) using the median expression values from the normalized counts matrix (log-transformed counts per million; with a prior count of 1) of the nine biological replicates from each condition. The function PCA from the R package AMOR (58) was employed with default parameters. The ellipses were defined using a multivariate t-distribution with a 70% confidence level. Code used for all analysis is found at https://github.com/ncolaian/mae.

DNA extraction for 16S sequencing. Plants were grown using MS medium (0.5X Murashige and Skoog basal medium with vitamins at pH 5.7 solidified with 1% bacto-agar) under the conditions and treatments described previously. Five roots were pooled and placed in 2.0 mL tubes with three sterile 4 mm glass beads. Samples were vortexed rigorously three times in sterile distilled water to remove agar particles and weakly associated microbes. Tubes containing the samples were stored at -80˚C until processing. Agar from each plate was collected in 30 mL syringes with a square of sterilized Miracloth (Millipore) at the bottom and stored at -20 °C until processing. Root samples were lyophilized for 48 hours using a Labconco freeze dry system and pulverized for 45 s in a FastPrep-24 Classic Instrument (MP Biomedicals). Agar containing syringes were thawed at room temperature and samples were squeezed gently through the Miracloth into 50 mL falcon tubes. Samples were centrifuged at max speed for 20 min and

most of the supernatant was discarded. The remaining 1-2 mL of supernatant, containing the pellet, was transferred into clean 2 mL tubes. Samples were centrifuged again, supernatant was removed, and pellets were stored at -80 °C until DNA extraction. DNA extractions were carried out on ground root and agar pellets using 96-well format MoBio PowerSoil Kit (MOBIO Laboratories; Qiagen) following the manufacturer's instruction.

16S rDNA Sequencing Preparation. We amplified the V3-V4 regions of the bacterial 16S rRNA gene using the primers 357F (5′-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGAGGCAGCAG-3′) and 806R (5′-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGACTACHVGGGTWTCTAAT-3′). Each PCR reaction was performed in triplicate. PCR conditions were as follows with the Kapa HiFi Hotstart readymix: 1.2 µL Kapa Buffer, 0.18 µL dNTPs, 0.3 µL DMSO, 0.12 µL Kapa HiFi Polymerase, 0.3 µL 10 µM 357F, 0.3 µL 10 µM 806R, 0.9 µL mixed plant rRNA gene-blocking peptide nucleic acids (PNAs; 1:1 mix of 10 µM plastid PNA and 10 µM mitochondrial PNA), 1.7 µL dH2O, 1 µL DNA; temperature cycling: 95°C for 5 min; 28 cycles of 98°C for 20 s; 78°C (PNA) for 15 s; 55°C for 15 s; 72°C for 1 min; 4°C until use. Triplicate PCR products were pooled and then diluted twice; 2 µL of PCR products into 18 µL of dH20 and 5 µL of Dilution 1 into 45 µL of dH2O. Following dilution, the PCR product was indexed using 16 indexed primers 357F (5'-AATGATACGGCGACCACCGAGATCTACACXXXXXXXXTCGTCGGCAGCGTC-3') and 24 indexed primers 806R (5'-CAAGCAGAAGACGGCATACGAGATXXXXXXXXGTCTCGTGGGCTCGG-3'). PCR conditions were as follows with the Kapa HiFi Hotstart readymix: 2 µL Kapa Buffer, 0.3 µL

dNTPs, 0.5 µL DMSO, 0.2 µL Kapa HiFi Polymerase, 0.5 µL 10 µM 357F index, 0.5 µL 10 µM 806R index, 0.15 µL PNAs; 1:1 mix of 100 µM plastid PNA and 100 µM mitochondrial PNA), 0.85 µL dH2O, 5 µL DNA; temperature cycling: 95°C for 5 min; 10 cycles of 98°C for 20 s; 78°C (PNA) for 15 s; 55°C for 15 s; 72°C for 1 min; 4°C until use. PCR products were purified using Sera-mag SpeedBeads (Fisher # 09-981-123) and quantified with a Qubit 2.0 fluorometer (Invitrogen) and then diluted to 6 pM for sequencing. Sequencing was performed on an Illumina MiSeq instrument using a 600- cycle V3 chemistry kit with 10% PhiX.

Analysis of the 16S sequencing data. Raw reads were trimmed with trimmomatic version 0.36 (44) using the parameters "HEADCROP:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15". Due to poor-quality base outputs in the beginnings and ends of reads we used headcrop and the leading/trailing trimming procedures. The sliding window command trims reads after the mean of 4 continuous base pairs reaches below a mean of 15. Once filtered, the resulting paired reads were analyzed using the DADA2 package version 1.14.1 (59) in R using the custom script code_for_16S_analysis.Rmd (https://github.com/ncolaian/MAE). Briefly, reads were filtered again by removing sequences containing uncalled bases, truncating reads after bases with a quality score of 2 and reads with no more than 3 expected errors using the FlilterAndTrim function in DADA2. We then learned the error rates for the forward and reverse reads separately with the learnErrors function. These error rates were then used to infer amplicon sequence variants (ASVs) with the function dada on the reverse and forward reads separately. After, we merged the forward and reverse strand results with the mergePairs function. The merged ASVs were used to construct an ASV sequencing table with only ASVs between 420 and 450 bp long, which reflect a majority of the reads from the V3 and V4 region, using the makeSequenceTable function. We then removed chimeras from the sequence table with the

removeBimeraDenova function. The resulting ASVs were then mapped to the known V3 and V4 regions of the isolates in our 35-member SynCom at 98% identity using vsearch's -usearch_global function (2.14.2). We only considered ASVs that mapped to SynCom members in all further 16S analysis.

Analysis of flg22 effect on the assembly of SynCom35. We performed our constrained analysis of principal coordinates (CAPs) similar to Finkel et al. (60). Briefly, the number of reads from ASVs that mapped to each isolate were summed together to reflect a single abundance value for each member of SynCom35. Relative abundances were calculated by dividing each count by the total number of reads mapping to the SynCom35 members in that sample. We performed a constrained analysis of principal coordinates (CAPs) on the root sample's relative abundance data using the capscale function from the vegan version 2.5-6 package in R. We used this function to create a square-rooted Bray-Curtis distance matrix with the vegdist function, and ordinate the matrix based on the model formula: relative_abundance ~ flg22 + experiment_batch. We then performed a PERMANOVA to determine the significance of the model and each of the model terms with 5000 permutations using the anova.cca function from vegan. We used the sum of variance explained calculated from the anova.cca function to determine the constrained and unconstrained variance explained.

Differential abundance of microbes with the addition of flg22 in root samples. We took the raw reads for each bacterial strain and used the DESeq2 version 1.26.0 (61) package in R to call differential abundant microbes between flg22- and mock-treated root samples. We created a DEseq object using the DESeqDataSetfromMatrix function and used the model – reads ~ Biological.Replicate + flg22. To perform the differential abundance analysis, we first estimated the normalization factors using the estimateSizeFactors function implemented in

DESeq2 using the poscounts method, which handles bacteria with 0 counts in some of the samples by calculating a modified geometric mean. We then estimated the dispersion estimates for the negative binomial distribution that will be used to call differential abundances with the estimateDispersions function in DESeq2. We then fit coefficients to each microbe indicating the change in abundance between treatments and tested these coefficients for significance using a Wald test implemented in the nbinomWaldTest function in DESeq2. We used an alpha value of 0.05 to assign significance.

Enrichment of members in root or agar samples. Each 16S experiment had a paired root and agar sample. We utilized this design to compare the relative abundance values for each suppressor (SynCom5-S1) and non-suppressor (SynCom-NS1) strain between the root and agar samples. We counted the number of times each strain was found at higher relative abundance values in either the agar or root sample. We did not compare relative abundance values if a strain was absent from both the root and agar samples in an experiment. The maximum number of experiments counted for each strain was 18, the minimum was 0 (for CL18 which was never found in any of our samples), and the median was 18. The counts (we called enrichments) for each strain were combined into the respective SynCom groupings for a total enrichment count of 68 for SynCom5-NS1 and 80 for SynCom5-S1. The agar and root enrichments were then tabulated (SynCom X enrichment) and a chi-square test was performed on the table with the chisq.test function in R.

Evaluation of the root acute response to flg22. We evaluated the ability of each member of SynCom35 to suppress the roots response to flg22 in the context of an acute exposure to the elicitor (i.e., 5 hours) using the procedures described by Millet et al. (25) (SI Appendix, Fig. 4.11D). For this, we used an Arabidopsis line carrying the pCYP71A12::GUS reporter

construct, which is activated by flg22 specifically in the root elongation zone (62). Approximately 10 to 15 bleach sterilized seeds (procedure described above) were inoculated in each well of a 12-well microtiter plate containing 1 mL MS medium (Murashige and Skoog basal medium with vitamins containing 0.5 g/L MES hydrate and 0.5% sucrose at pH 5.7). The plate was kept in a growth chamber under 16-hour light/8-hour dark regime at 21oC day/18oC night for seven days. At the end of the seventh day, the medium was removed with a pipette and replaced by 1 mL MS medium only ('no bacteria' control) or 1 mL MS containing a single bacterial strain at OD600nm of 0.002. The plate was returned to the growth chamber and incubated overnight (~14 hours) to allow bacterial colonization of the roots. The elicitor, flg22, was then added to the medium at a final concentration of 100 nM. Seedlings treated with Pseudomonas simiae WCS417 (overnight incubation) or with 10 µM MeJA (added along with flg22) were used as positive controls for the suppression of the flg22 response (62). Plates were kept in the growth chamber for 5 hours after flg22 addition and then submitted to the GUS histochemical assay.

GUS histochemical assays. After the treatment with bacteria and/or flg22, seedlings were washed twice with 50 mM sodium phosphate buffer, pH 7. Then, 1 mL of freshly prepared GUS substrate solution (50 mM sodium phosphate, pH 7; 10 mM EDTA; 0.5 mM K4[Fe(CN)6]; 0.5 mM K3[Fe(CN)6]; 0.5 mM X-Gluc; and 0.01% Silwet L-77) was added to each well of the plate, which was subsequently incubated in the dark for 4 hours at 37oC. GUS substrate solution was then replaced by 1 mL of a 3:1 ethanol:acetic acid solution and left at 4oC overnight (~14 hours). Before imaging, this solution was replaced with 1 mL 95% ethanol. Roots were visualized using a Leica M205FA stereoscope coupled to a Leica DFC310FX camera.

Root immune suppression by bacterial culture. We evaluated if culture filtrates of suppressor strains are sufficient to prevent the root response to flg22 following the procedures described by Yu et al. (63) (SI Appendix, Fig. 4.13A). Briefly, Arabidopsis seedlings were surface sterilized and grown for 10 days in a 12-well microtiter plate containing 1 mL MS medium as described above. After this, the plant growth medium was filtered using a 0.22 μM Millipore syringe filter and stored at -20°C until use. Bacteria were grown in 2xYT medium, washed in 10 mM MgCl2 and then inoculated in a 12-well plate containing the plant exudate filtrate at a final OD_{600nm} of 0.002. After incubating the plate for 22 hours in a growth chamber, cultures were passed through a 0.22 μM Millipore syringe filter, resulting in a bacteria filtrate. This filtrate was then supplied to 10-day-old pCYP71A12::GUS Arabidopsis seedlings for 1.5 hours, and 100 nM of flg22 (or a mock control) was then added. After five hours of flg22 treatment, seedlings were submitted to the GUS histochemical assay as described above. Plants growing in the absence of any bacteria or in the culture filtrate of the non-suppressor strain *Ochrobactrum* sp. MF370 were used as controls. The experiment also included *Pseudomonas simiae* strain WCS417, whose acidic culture filtrate suppresses the plant response to flg22 (63). To contrast our suppressor strains to *P. simiae* WCS417, we measured the pH of all bacterial filtrates using an Accumet ab15 pH meter (Fisher Scientific).

Phylogenetic tree construction and annotation. To construct a phylogenetic tree of the 35- member SynCom strains, we trimmed a previously constructed tree (60) that contained a total of 185 strains, including all members of the 35-member SynCom, using the drop.tip function from the ape package version 5.3 in R (64). This phylogenetic tree was based on 47 single copy genes that are present in every isolate of the larger 185-member collection. Single gene alignments were performed using MAFFT (65) and low-quality columns were filtered with

trimAl (66). Gene alignments were then concatenated into a single super-alignment and used in the phylogenetic inference using FastTree version 2.1 with the WAG evolution model (67). The resulting tree was annotated with the results from five of our experiments that evaluated the ability of each strain to suppress/activate the plant immune responses: (I) Root-growth inhibition (RGI) triggered by flg22 (Fig. 4.2A); (II) Expression of the flg22 regulon (Fig. 2B); (III) Expression of the flg22 core gene set (SI Appendix, Fig. 4.11C); (IV) Interference with the plant response to acute exposure to flg22 (SI Appendix, Fig. 4.11D); (V) Expression of specific sectors of the flg22 regulon (Clusters M3, M4 and M5; Fig. 4.2C).

Quantification of bacterial colonization of plant roots. We evaluated the bacterial levels in plants colonized by communities of suppressor or non-suppressor strains in two independent assays. The first assay (Fig. 4.3C) was based on a community of five suppressors (SynCom5-S1) and five taxonomically matching non-suppressors (SynCom5-NS1). In this experiment, Arabidopsis Col-0 were grown axenically in Johnson medium for seven days (as described above for the SynCom35 experiments) and then transferred to new plates containing either SynCom5-S1 or SynCom5-NS1 (105 cells/mL each strain; prepared as described above for SynCom35). Plants transferred to medium without bacteria were used as controls. Plates were kept in a growth chamber under a 9-hour light/15-hour dark regime at 21oC day/18oC night. The root systems of approximately 5 plants per plate were harvested after 12 days for bacteria counting. The experiment was repeated 5 independent times, each with 3 biological replicates per condition. The second assay (SI Appendix, Fig. 4.13) utilized a slightly modified community of five suppressors (SynCom5-S2) and five different, not taxonomically matching, non-suppressors (SynCom5-NS2). In this experiment, Arabidopsis Col-0 plants were germinated in 0.5xMS medium in the presence of SynComs (5x105 cells/mL each strain) or no bacteria and

kept in a growth chamber under a 9-hour light/15-hour dark regime at 21oC day/18oC night. The root system of 3-6 seedlings were harvested after 16 days. Strains that comprise all SynComs are listed in SI Appendix, Dataset 1.

To isolate and quantify plant-colonizing bacteria, roots were harvested, rinsed and vortexed vigorously three times with sterile 10 mM MgCl2 to remove agar particles and weakly-associated microbes. Plant material was weighed and then homogenized for 45s in a FastPrep-24 Classic Instrument (MP Biomedicals) in 2 mL tubes containing three 4 mm glass beads and 400 µL 10 mM MgCl2. Homogenized samples were brought to 1 mL by adding 600 µL 10 mM MgCl2, then submitted to a serial dilution. Four microliters of each dilution were inoculated onto LB agar plates (Luria-Bertani; 10g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl and 15 g/L agar). Colony forming units were counted after 1-3 days incubation at 28oC and used to determine the original bacterial abundance per gram of root tissue based on the serial dilution used for counting.

Evaluation of the Arabidopsis transcriptional response to SynComs of suppressors and non-suppressors. We evaluated the transcriptomes of plants grown in the presence of five suppressors (SynCom5-S1) or five taxonomically matching non-suppressors (SynCom5-NS1) (SI Appendix, Fig. 4.14B-E). This analysis is part of the experiment described above for the quantification of bacterial growth (shown in Fig. 4.3C). Briefly, seven-day-old Col-0 seedlings were transferred to plates containing no bacteria, SynCom5-S1 or SynCom5-NS1 (105 cells/mL each strain; SI Appendix, Dataset 1). No exogenous flg22 was added. Whole seedlings were harvested after twelve days and used for RNA extraction and sequencing as described above. The experiment included six biological replicates per conditions and was repeated two independent times (each with 3 biological replicates). The processing of RNA-seq reads,

identification of differentially expressed genes and gene ontology enrichment analyses were performed as described for the other RNA-seq experiments described above. We compared the transcriptomes of plants grown with each SynCom against the 'no bacteria' control using the model: expression ~ SynCom + experiment, where "SynCom" refers to the presence of bacteria (no bacteria, SynCom5-S1 or SynCom5-NS1) and "experiment" refers to each of the two experimental batches. SI Appendix, Dataset 6 presents the metadata information of the experiment, the lists of differentially expressed genes and the complete results of the enrichment analyses.

Evaluation of the growth of commensals in the presence of SynComs. We measured if the root colonization capacity of commensal bacteria (*Ochrobactrum* sp. MF370 or *Pseudomonas viridiflava* OTU5 p5.e6) is influenced by communities of suppressors (SynCom5-S2) or nonsuppressors (SynCom5-NS2) (Fig. 4.4 and SI Appendix, Dataset 1). For this, approximately 500 Col0 seeds were sterilized by shaking in 4.25% sodium hypochlorite for 4 min. The sodium hypochlorite was removed, and the seeds were washed four times in sterile water and then resuspended in 0.1% agar. Seeds were vernalized at 4°C for 2-3 days. Seeds were germinated on 0.5X MS medium containing either the suppressor SynCom (SynCom5-S2), the non-suppressor SynCom (SynCom5-NS2) or no bacteria. Plates were kept in a growth chamber under a 9-hour light/15-hour dark regime at 21°C day/18°C night. All bacteria strains were cultured from a single colony and grown in 3 mL of LB medium for 24 h at 28° C and 250 rpm. Cells were centrifuged for 6 min at 4,342 x g and washed three times with 10 mM MgCl2. Media containing SynComs were prepared by embedding each strain at an OD600nm of 0.0005. After two weeks, plants were flooded for 5 min with 7 mL of *Ochrobactrum* sp. MF370 or *Pseudomonas viridiflava* OTU5 strain p5.e6 containing pBBR1MCS-5 at an OD600nm of

0.0001. Seedlings were then transferred to new 0.5X MS agar plates without bacteria. After 30 min, the flooding and transfer step was repeated with the strain in 10 mM MgCl2 containing 0.005% silwet. After 48 h, roots from 3-6 seedlings per replicate were harvested, weighed, and ground for 45 s in a FastPrep-24 Classic Instrument (MP Biomedicals) in a 2 mL microcentrifuge tube containing 400 µL of 10 mM MgCl2 and three 4 mm glass beads (Fisher, Cat. No. 11-312B). Homogenized samples were brought up to 1 mL by adding 600 µL 10 mM MgCl2, then submitted to a serial dilution. Four microliters of each dilution were inoculated onto LB agar plates (Luria-Bertani; 10g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl and 15 g/L agar) containing 25 µg/mL gentamycin or 50 µg/mL kanamycin for antibiotic selection of OTU5 and MF370, respectively. Members of SynCom5-S2 and SynCom5-NS2 were susceptible to these antibiotics. Experiments were performed with three biological replicates and repeated three to four times. All experiments were analyzed together using an ANOVA that controlled for a significant batch effect using the model: log.CFU ~ treatment + batch. We identified truly significant groups with a Tukey test and an alpha of 0.05 using the HSD.test function from the agricolae package in R (68).

Type-3 secretion system deletion in MF79. Two unmarked Type-3 Secretion System (T3SS) deletion mutants of Dyella japonicum UNC79MFTsu3.2 (henceforth MF79) were constructed by two-step allelic exchange to test the involvement of the T3SS in suppression of the flg22 response observed by wild-type MF79. The first strain, MF79 *ΔsctUTSR*, has the genes for the inner membrane export apparatus of the T3SS deleted (*sctUTSR*, IMG database gene ID 2558297553- 2558297556). The second strain, MF79 ΔT3SS, has a larger deletion comprising the full T3SS (sctUTSRVCDJLN, IMG database gene ID 2558297553-2558297569). Knockout vectors pJMC151 and pJMC152 for construction of strains MF79 Δ*sctUTSR* and MF79 ΔT3SS,

respectively, were designed based on a genetic system developed for *Burkholderia* spp. and the suicide vector pMo130 (69). All PCR steps were performed with Q5 High-Fidelity DNA Polymerase (New England Biolabs) including the optional GC enhancer. The pMo130 vector backbone was amplified using primers JMC203-JMC204, cleaned-up, and treated with DpnI (New England Biolabs). 800-1000bp regions for homologous recombination flanking genes to be deleted were amplified (pJMC151/152 5': primers JMC403-JMC404, pJMC151 3': primers JMC405-JMC406, pJMC152 3': primers JMC407-JMC408). Vectors pJMC151 and pJMC152 were assembled using HiFi DNA Assembly Mastermix (New England Biolabs), transformed into NEB 5-alpha chemically competent E. coli (New England Biolabs) and selected on LB agar with 50 μg/mL kanamycin at 37°C. Plasmid DNA was isolated from clones using the ZR Plasmid Miniprep Classic Kit (Zymo Research) and sequenced by Sanger sequencing. Sequencing-confirmed pJMC151 and pJMC152 were transformed into biparental mating E. coli strain WM3064 and selected on LB agar plates containing 50 μg/mL kanamycin and 0.3 mM diaminopimelic acid (DAP) at 37°C.

Biparental mating was performed by growing the bacteria overnight: *Dyella japonicum* MF79 in 2xYT medium at 28°C and *E. coli* WM3064 containing either pJMC151 or pJMC152 in LB with 50 μg/mL kanamycin and 0.3 mM DAP at 37°C. Bacteria were washed three times with 2xYT medium and resuspended in 1/10 of the original volume and mixed in equal proportion donor to recipient and plated on LB agar plates containing 0.3 mM DAP and grown overnight at 28°C. The following day, exconjugants of MF79 were selected by streaking on LB agar plates containing 50 μg/mL kanamycin and lacking DAP and grown at 28°C. Several putative first crossover strains were grown from individual colonies in 2xYT medium containing 50 μg/mL kanamycin and screened using primers outside the regions of homologous

recombination (JMC409, JMC410, JMC411) and primers on the pMo130 suicide vector backbone (JMC321, JMC322) to confirm insertion of knockout vectors in the MF79 chromosome at the appropriate location. Positive first crossovers were resolved by passaging the strains once on 2xYT lacking antibiotics, then passaging on media containing 10 g/L tryptone, 5 g/L yeast extract, 100 g/L sucrose, and 1mM IPTG, and finally plating on the same medium containing 1.5% agar. The resulting strains were screened for gene deletion using primers JMC409-JMC410 for strain MF79 *ΔsctUTSR* and primers JMC409-JMC411 for strain MF79 *ΔT3SS*. Putative positive strains were plate-purified by streaking two additional times on LB agar plates. The gene deletions in the final strains were confirmed again using PCR as described above. Purity of the final deletion strains were confirmed by performing PCR using forward primer JMC409 and reverse primer JMC426 inside the deleted gene sctU, which resulted in a 1.2 kb product for wild-type MF79 and no product in the deletion strains. Primers are listed in SI Appendix, Dataset 7 and the confirmatory DNA gel is shown in SI Appendix, Fig. 4.15A. Growth of the knockout mutants was compared to wild-type MF79 by growing the strains in Minimal Medium A (70) with 10 mM sucrose and 0.1% tryptone in an Infinite M200 Pro plate reader (Tecan) at 28°C while measuring OD_{600nm}.

MF79 Transposon insertion library screening. Arabidopsis pCYP71A12::GUS seeds were surface sterilized and stratified as described above for 24-48 hours in sterile distilled water, then arrayed in 96-well plates, with 1-3 seeds per well. 80 µL of MS medium was added and seedlings were grown at 16-hour light/8-hour dark regime at 21°C day/18°C night for 7 days.

The RB-TnSeq library for MF79 (71) was plated on LB agar plates containing 50 µg/mL kanamycin and grown at 28°C. Approximately 4500 individual colonies were placed into single wells of 96- deep well plates containing 2xYT medium with 50 µg/mL kanamycin and grown at

28°C. The bacteria were washed 3 times with 10 mM $MgCl_2$, resuspended in MS medium, and then 1 µL was added to each well containing Arabidopsis *pCYP71A12::GUS* seedlings in 80 µL MS. To store the arrayed bacteria, a final concentration of 20% glycerol was added to each well of washed bacteria and the library was frozen at -80°C. In total, 49 plates were screened comprising approximately 4500 individual transposon insertion strains.

Approximately 16 hours after adding bacteria to the seedlings, 1 µM flg22 peptide was added to the wells. After 5 hours of flg22 treatment, seedlings were submitted to the GUS histochemical assay as described above. Plates were manually inspected for any wells containing plants with blue root tips indicating a failure to suppress the flg22 induction of *pCYP71A12*-driven GUS expression. All putatively positive strains were regrown by streaking from the 96-well plate glycerol stock of the arrayed library to LB plates containing 50 µg/mL kanamycin. Colonies were picked and grown 2xYT medium containing 50 µg/mL kanamycin, washed three times with 10mM $MgCl_2$, and assayed again using the GUS histochemical assay in 48-well plates with 100 nM flg22 peptide to confirm their ability to suppress the flg22 induced GUS expression. In this assay, bacteria were inoculated at an OD_{600nm} of 0.002. Six strains from the mutant library screened in this confirmatory assay were unable to suppress the flg22 induced GUS expression and were subjected to transposon insertion mapping. Growth of the selected Tn mutants were compared to wild-type MF79 by growing the strains in Minimal Medium A (70) with 10 mM sucrose and 0.1% tryptone in an Infinite M200 Pro plate reader (Tecan) at 28°C while measuring OD_{600nm}.

Transposon insertion mapping. An arbitrary PCR approach was used to map the transposon insertions in the six strains that lost the ability to suppress the plant response to flg22. PCR steps were performed with Q5 High-Fidelity DNA Polymerase (New England Biolabs)

using the optional GC enhancer according to the manufacturer instructions. A first PCR was performed using a transposon kanamycin marker specific primer (JMC525) and arbitrary primer ARB6 (72) with an annealing temperature of 35°C and extension time of 90s for 30 cycles. Following this, a second round of PCR was performed on the product using a primer nested inside the kanamycin resistance gene of the expected product (JMC567) and primer ARB2 (72) with annealing temperature of 71°C and extension time of 90s for 30 cycles. The resulting PCR product was cleaned up and Sanger sequenced using primer JMC567. BLASTn was used to identify the MF79 genomic sequence adjacent to the transposon insertion and determine the Tn insertion location and, thus, the gene disrupted by the insertion. Primers are listed in SI Appendix, Dataset 7.

Suppression of the flg22 response by D. japonica MF79 cell-free supernatant. D. japonica MF79 wild-type and the gspD::Tn and gspE::Tn mutant strains were grown to stationary phase in 2xYT medium. Each strain was washed three times with 10 mM MgCl2 and inoculated at OD600nm of 0.02 into 1xMS medium containing 5 g/L sucrose and 0.1% (w/v) casamino acids (Bacto). Cultures were grown at 21°C with 250 rpm shaking for 18 hours. To remove the bacteria, cultures were centrifuged at 4000 x g for 15 min and the supernatant was filtered through as 0.2 μm filter to prepare cell-free supernatant. To evaluate the potential size of the active suppressive molecule(s), supernatant from the MF79 wild type was passed through a 10,000 MWCO Amicon Ultra-15 spin concentrator (Millipore). The 10 kDa flow-through is the sample that passed through the concentrator. The 10 kDa retentate sample was retained in the concentrator and brought to its original volume with 1xMS medium. Both the 10 kDa retentate and flow-through were sterile filtered prior to use.

pCYP71A12::GUS seedlings were grown in 24-well plates with 500 µL 1xMS medium with sucrose. This growth medium was removed from 7-day-old seedlings and replaced with the sterile cell-free supernatant, 10 kDa retentate, or 10 kDa flow-through samples from MF79 wild-type, *gspD::Tn*, and *gspE::Tn* or a blank medium control. Then, 1 µM flg22 was added to the wells for 5 hours, following which GUS histochemical staining was performed as described above.

Non-suppressor MF370 growth with MF79 10 kDa suppressive fraction. We evaluated if suppression of the flg22 response by *D. japonica* MF79 supernatant enhanced the growth of a nonsuppressor commensal strain (Fig. 4.5E). Arabidopsis *pCYP71A12::GUS* seeds were surface sterilized as described above, arrayed in 24-well plates with 5-6 seeds per well, and grown in 500 µL 1xMS medium containing sucrose at 16-hour light/8-hour dark regime at 21°C day/18°C night for 7 days. Then, the wells and seedlings were washed 3 times with 750 µL 0.5x MS lacking sucrose.

Cell-free supernatant from *D. japonica* MF79 wild-type, *gspD::Tn*, and *gspE::Tn* were prepared as described above. To remove sucrose and casamino acids, the cell free supernatant for each was concentrated in a 10,000 MWCO concentrator and exchanged into 0.5xMS lacking sucrose by adding this medium and concentrating three times. The final exchanged 10 kDa retentate was brought to the original volume in 0.5xMS lacking sucrose. 500 µL of 10 kDa retentate exchanged into 0.5xMS or medium alone as a control was added to the wells of the washed 7-day old seedlings and they were returned to the growth chamber. Approximately 18 hours later, 1 µM flg22 or mock treatment was added to wells for 5 hours. The non-suppressor strain *Ochrobactrum* sp. MF370 was grown to stationary phase in 2xYT medium containing 100 µg/mL ampicillin and washed three times in 10 mM MgCl2. *Ochrobactrum* sp. MF370 was

diluted in 0.5xMS and added to the wells at a final OD600nm of 0.0002. Plates were returned to the growth chamber for 2 days. To assess the colonization by *Ochrobactrum* sp. MF370, plants were harvested as described above for the plant colonization competition assay and colony forming units of *Ochrobactrum* sp. MF370 per g plant fresh weight were determined on LB plates containing 100 µg/mL ampicillin.

Statistical information. All statistical tests were performed using R software. All boxplots display the median, 25th and 75th quantiles (box), and the smallest value, largest value, or 1.5*IQR below or above the 25th and 75th quantiles (whiskers). All statistical tests used an alpha value of 0.05 to assign significance unless otherwise stated. When possible, all data points are presented in the figures from all replications performed unless otherwise stated. ANOVAs were performed as a two-way ANOVA to control for experimental biases when analyzing multiple experiments at once. Any time multiple comparisons were made, we corrected our comparisons using a Tukey test. When many statistical tests were performed on the same data and many p-values were obtained we performed Benjamini and Hochberg p-value correction method (FDR) to control for false discovery rate. All statistical analysis methods are described in the figure legends. Code for statistical tests can be found at GitHub (https://github.com/ncolaian/MAE).

Data and code availability. All data generated in this project is publicly available. Raw sequences of 16S amplicon sequencing are available at the NCBI Short Read Archive (SRA) under the accession number PRJNA657936. The metadata file, read count table, relative abundance matrix and DESeq results of this experiment are provided in SI Appendix, Dataset 2. Raw sequences and read count matrices from RNA-seq experiments are available at the NCBI Gene Expression Omnibus (GEO) under the accession number GSE156426. The metadata file,

edgeR results, list of heatmap cluster members and raw results from Gene Ontology (GO) enrichment analyses are provided in SI Appendix, Datasets 3 and 4. Scripts used for plotting and data analysis are available at GitHub (https://github.com/ncolaian/MAE).

REFERENCES

1. C. R. Fitzpatrick, *et al.*, The Plant Microbiome: From Ecology to Reductionism and Beyond. *Annu. Rev. Microbiol.* **74**, annurev-micro-022620-014327 (2020).

2. P. J. P. Teixeira, N. R. Colaianni, C. R. Fitzpatrick, J. L. Dangl, Beyond pathogens: microbiota interactions with the plant immune system. *Curr. Opin. Microbiol.* **49**, 7–17 (2019).

3. K. Yu, C. M. J. Pieterse, P. A. H. M. Bakker, R. L. Berendsen, Beneficial microbes going underground of root immunity. *Plant Cell Environ.* **42**, 2860–2870 (2019).

4. S. Hacquard, S. Spaepen, R. Garrido-Oter, P. Schulze-Lefert, Interplay Between Innate Immunity and the Plant Microbiota. *Annu Rev Phytopathol* **55**, 565–589 (2017).

5. S. L. Lebeis, *et al.*, PLANT MICROBIOME. Salicylic acid modulates colonization of the root microbiome by specific bacterial taxa. *Science* **349**, 860–864 (2015).

6. J. M. Kniskern, M. B. Traw, J. Bergelson, Salicylic Acid and Jasmonic Acid Signaling Defense Pathways Reduce Natural Bacterial Diversity on *Arabidopsis thaliana*. *MPMI* **20**, 1512–1522 (2007).

7. T. Chen, *et al.*, A plant genetic network for preventing dysbiosis in the phyllosphere. *Nature* **580**, 653–657 (2020).

8. F. Zhou, *et al.*, Co-incidence of Damage and Microbial Patterns Controls Localized Immune Responses in Roots. *Cell* **180**, 440-453.e18 (2020).

9. A. Emonet, *et al.*, Spatially Restricted Immune Responses Allow for Root Meristematic Activity During Bacterial Colonisation. *bioRxiv* (2020).

10. L. da Cunha, M.-V. Sreerekha, D. Mackey, Defense suppression by virulence effectors of bacterial phytopathogens. *Current Opinion in Plant Biology* **10**, 349–357 (2007).

11. N. Geva-Zatorsky, *et al.*, Mining the Human Gut Microbiota for Immunomodulatory Organisms. *Cell* **168**, 928-943.e11 (2017).

12. K. Yu, *et al.*, Rhizosphere-Associated *Pseudomonas* Suppress Local Root Immune Responses by Gluconic Acid-Mediated Lowering of Environmental pH. *Current Biology* **29**, 3913-3920.e4 (2019).

13. R. Garrido-Oter, *et al.*, Modular Traits of the Rhizobiales Root Microbiota and Their Evolutionary Relationship with Symbiotic Rhizobia. *Cell Host Microbe* **24**, 155-167.e5 (2018).

14. Z. Liu, *et al.*, A Genome-Wide Screen Identifies Genes in Rhizosphere-Associated Pseudomonas Required to Evade Plant Defenses. *mBio* **9** (2018).

15. I. A. Stringlis, *et al.*, Root transcriptional dynamics induced by beneficial rhizobacteria and microbial immune elicitors reveal signatures of adaptation to mutualists. *Plant J.* **93**, 166–180 (2018).

16. J. M. Plett, *et al.*, Effector MiSSP7 of the mutualistic fungus *Laccaria bicolor* stabilizes the Populus JAZ6 protein and represses jasmonic acid (JA) responsive genes. *Proceedings of the National Academy of Sciences* **111**, 8299–8304 (2014).

17. Y. Liang, *et al.*, Nonlegumes respond to rhizobial Nod factors by suppressing the innate immune response. *Science* **341**, 1384–1387 (2013).

18. V. Lakshmanan, *et al.*, Microbe-associated molecular patterns-triggered root responses mediate beneficial rhizobacterial recruitment in Arabidopsis. *Plant Physiol.* **160**, 1642–1661 (2012).

19. Y. A. Millet, *et al.*, Innate immune responses activated in Arabidopsis roots by microbe-associated molecular patterns. *Plant Cell* **22**, 973–990 (2010).

20. G. Rallapalli, *et al.*, EXPRSS: an Illumina based high-throughput expression-profiling method to reveal transcriptional dynamics. *BMC Genomics* **15**, 341 (2014).

21. C. Denoux, *et al.*, Activation of defense response pathways by OGs and Flg22 elicitors in Arabidopsis seedlings. *Mol Plant* **1**, 423–445 (2008).

22. G. Castrillo, *et al.*, Root microbiota drive direct integration of phosphate stress and immunity. *Nature* **543**, 513–518 (2017).

23. P. J. Rushton, I. E. Somssich, P. Ringler, Q. J. Shen, WRKY transcription factors. *Trends in Plant Science* **15**, 247–258 (2010).

24. D. Couto, C. Zipfel, Regulation of pattern recognition receptor signalling in plants. *Nat Rev Immunol* **16**, 537–552 (2016).

25. T. Gigolashvili, *et al.*, The transcription factor HIG1/MYB51 regulates indolic glucosinolate biosynthesis in *Arabidopsis thaliana*. *Plant J.* **50**, 886–901 (2007).

26. A. Pascale, S. Proietti, I. S. Pantelides, I. A. Stringlis, Modulation of the Root Microbiome by Plant Molecules: The Basis for Targeted Disease Suppression and Plant Growth Promotion. *Front Plant Sci* **10**, 1741 (2019).

27. K.-W. Ma, *et al.*, Coordination of microbe-host homeostasis via a crosstalk with plant innate immunity. *Research Square (pre-print)* (2020).

28. T. M. Winkelmüller, *et al.*, Gene expression evolution in pattern-triggered immunity within *Arabidopsis thaliana* and across Brassicaceae species. *bioRxiv* (2020).

29. L. Gómez-Gómez, G. Felix, T. Boller, A single locus determines sensitivity to bacterial flagellin in *Arabidopsis thaliana*: Sensitivity of *A. thaliana* to bacterial flagellin. *The Plant Journal* **18**, 277–284 (1999).

30. O. M. Finkel, *et al.*, A single bacterial genus maintains root growth in a complex microbiome. *Nature* (2020) https:/doi.org/10.1038/s41586-020-2778-7.

31. T. L. Karasov, *et al.*, *Arabidopsis thaliana* and *Pseudomonas* Pathogens Exhibit Stable Associations over Evolutionary Timescales. *Cell Host Microbe* **24**, 168-179.e4 (2018).

32. W. Deng, *et al.*, Assembly, structure, function and regulation of type III secretion systems. *Nat Rev Microbiol* **15**, 323–337 (2017).

33. M. N. Price, *et al.*, Mutant phenotypes for thousands of bacterial genes of unknown function. *Nature* **557**, 503–509 (2018).

34. V. Lakshmanan, R. Castaneda, T. Rudrappa, H. P. Bais, Root transcriptome analysis of *Arabidopsis thaliana* exposed to beneficial *Bacillus subtilis* FB17 rhizobacteria revealed genes for bacterial recruitment and plant defense independent of malate efflux. *Planta* **238**, 657–668 (2013).

35. C. Vogel, N. Bodenhausen, W. Gruissem, J. A. Vorholt, The Arabidopsis leaf transcriptome reveals distinct but also overlapping responses to colonization by phyllosphere commensals and pathogen infection with impact on plant health. *New Phytol.* **212**, 192–207 (2016).

36. B. Gourion, F. Berrabah, P. Ratet, G. Stacey, Rhizobium–legume symbioses: the crucial role of plant immunity. *Trends in Plant Science* **20**, 186–194 (2015).

37. N. P. Cianciotto, R. C. White, Expanding Role of Type II Secretion in Bacterial Pathogenesis and Beyond. *Infect. Immun.* **85** (2017).

38. C. Denoux, et al., Activation of defense response pathways by OGs and Flg22 elicitors in Arabidopsis seedlings. *Mol Plant* 1, 423–445 (2008).

39. G. Rallapalli, et al., EXPRSS: an Illumina based high-throughput expression-profiling method to reveal transcriptional dynamics. *BMC Genomics* 15, 341 (2014).

40. B. Li, et al., Phosphorylation of trihelix transcriptional repressor ASR3 by MAP KINASE4 negatively regulates Arabidopsis immunity. *Plant Cell* 27, 839–856 (2015).

38. G. Castrillo, et al., Root microbiota drive direct integration of phosphate stress and immunity. *Nature* 543, 513–518 (2017).

39. C. A. Schneider, W. S. Rasband, K. W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9, 671–675 (2012).

40. H. Wickham, ggplot2: elegant graphics for data analysis (springer, 2016).

41. Z. Gu, R. Eils, M. Schlesner, Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 32, 2847–2849 (2016).

42. S. S. Abby, B. Néron, H. Ménager, M. Touchon, E. P. C. Rocha, MacSyFinder: a program to mine genomes for molecular systems with an application to CRISPR-Cas systems. *PLoS ONE* 9, e110726 (2014).

43. J. Logemann, J. Schell, L. Willmitzer, Improved method for the isolation of RNA from plant tissues. *Anal. Biochem*. 163, 16–20 (1987).

44. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120 (2014).

45. D. Kim, B. Langmead, S. L. Salzberg, HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360 (2015).

46. Y. Liao, G. K. Smyth, W. Shi, The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res*. 41, e108 (2013).

47. P. Ewels, M. Magnusson, S. Lundin, M. Käller, MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32, 3047–3048 (2016).

48. D. J. McCarthy, Y. Chen, G. K. Smyth, Differential expression analysis of multifactor RNASeq experiments with respect to biological variation. *Nucleic Acids Res*. 40, 4288–4297 (2012).

49. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140 (2010).

50. M. D. Robinson, A. Oshlack, A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol*. 11, R25 (2010).

51. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)* 57, 289–300 (1995).

52. H. Heberle, G. V. Meirelles, F. R. da Silva, G. P. Telles, R. Minghim, InteractiVenn: a webbased tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics* 16, 169 (2015).

53. T. Wei, V. Simko, R package "corrplot": Visualization of a Correlation Matrix (2017).

54. J. Oksanen, et al., vegan: Community Ecology Package (2019).

55. G. R. Warnes, et al., gplots: Various R Programming Tools for Plotting Data (2020).

56. G. Yu, L.-G. Wang, Y. Han, Q.-Y. He, clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16, 284–287 (2012).

57. X. Yi, Z. Du, Z. Su, PlantGSEA: a gene set enrichment analysis toolkit for plant community. *Nucleic Acids Res.* 41, W98-103 (2013).

58. S. Herrera-Paredes, AMOR: Abundance Matrix Operations in R (2018).

59. B. J. Callahan, et al., DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13, 581–583 (2016).

60. O. M. Finkel, et al., A single bacterial genus maintains root growth in a complex microbiome. *Nature* (2020) https:/doi.org/10.1038/s41586-020-2778-7.

61. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550 (2014).

62. Y. A. Millet, et al., Innate immune responses activated in Arabidopsis roots by microbeassociated molecular patterns. *Plant Cell* 22, 973–990 (2010).

63. K. Yu, et al., Rhizosphere-Associated Pseudomonas Suppress Local Root Immune Responses by Gluconic Acid-Mediated Lowering of Environmental pH. *Current Biology* 29, 3913-3920.e4 (2019).

64. E. Paradis, K. Schliep, ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35, 526–528 (2019).

65. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780 (2013).

66. S. Capella-Gutiérrez, J. M. Silla-Martínez, T. Gabaldón, trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973 (2009).

67. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS ONE* 5, e9490 (2010).

68. F. de Mendiburu, agricolae: Statistical Procedures for Agricultural Research (2020).

69. M. A. Hamad, S. L. Zajdowicz, R. K. Holmes, M. I. Voskuil, An allelic exchange system for compliant genetic manipulation of the select agents Burkholderia pseudomallei and Burkholderia mallei. *Gene* 430, 123–131 (2009).

70. F. Ausubel, et al., Current Protocols in Molecular Biology (Wiley Interscience, New York, 1996).

71. M. N. Price, et al., Mutant phenotypes for thousands of bacterial genes of unknown function. *Nature* 557, 503–509 (2018).

72. G. A. O'Toole, et al., Genetic approaches to study of biofilms. *Meth. Enzymol*. 310, 91–109 (1999).

73. D. S. Lundberg, et al., Defining the core Arabidopsis thaliana root microbiome. *Nature* 488, 86–90 (2012).

74. K.-W. Ma, et al., Coordination of microbe-host homeostasis via a crosstalk with plant innate immunity. *Research Square* (pre-print) (2020). 72

CHAPTER 5: CONCLUSIONS AND PERSPECTIVES

Section 5.1 Conclusions and perspectives text

Collectively, this work provides a greater understanding of the interplay between host immune systems and host-associated commensal communities. The model plant, Arabidopsis, and the model MAMP, flg22, provide a glimpse into how microbial sensing receptors located at the extracellular surface of plant cells respond to microbial derived molecules. Since their identification in pathogens, the numerous studies and models proposed have focused on the defense against pathogens. However, the work described in this dissertation shows that the response to flg22 is more complex than previously believed and may play a larger role in commensal community formation.

Host immune responses exert evolutionary pressure on their associated microbiota that can influence the evolution of immunogenic MAMPs to non-immunogenic and antagonistic flg22 variants. In the context of a commensal community, diverse flg22 variants can antagonize, modulate, and uniquely activate plant responses. Flg22 sequences and responses are predictive of the class level identity of the microbe that causes the response, potentially indicating FLS2 functions as a community sensor. Many flg22 variants found in commensal microbiota are immunogenic. To potentially overcome the presence of immunogenic flg22, diverse commensal community members employ non-overlapping mechanisms to suppress/modulate the response to immunogenic flg22. We find that the suppressor strains from a synthetic community do not use a T3SS, instead at least one strain requires the general T2SS to suppress flg22 induced responses. Lastly, bacteria with immunomodulatory function have an advantage when colonizing the plant

in the community context and can help non-suppressing strains colonize plants.

In my dissertation, I argued that the traditional defense-oriented view of MTI is incomplete, and, as a result, its role in commensal community assembly and maintenance misunderstood. Our data confirm that immunogenic MAMP variants are detrimental to host colonization given the significant evolutionary pressure to evade immunity identified in *Pseudomonas* isolates. However, immunogenic variants are clearly abundant in commensal community members, indicating that some level of immunogenic MAMPs is tolerated. Thus, the presence of immunogenic MAMPs alone cannot function solely as a defense inducing signal. So how might it work?

I hypothesize that FLS2 works as a molecular tabulator that monitors the community for dysbiosis. Together plant PRRs take a census of MAMPs among its associated microbiota in order to respond accordingly. In this model, microbial diversity is promoted because of its ability to buffer the presence of immunogenic MAMPs with antagonists, immunomodulatory molecules, and suppression. That is, the heterogeneity of MAMPs produced by a robust, diverse community dilutes and diffuses the signal from the subset of pathogenic bacteria as long as their relative abundance remains small. The activation of defense through PRRs occurs when a pathogen strain producing immunogenic MAMPs dominates the commensal community. Pathogen infection generally leads to a reduction in bacterial diversity and dysbiosis. Thus, immunogenic MAMPs produced by the community are no longer buffered by MAMP and suppression mechanism diversity, leading to defense activation. In addition, the sensitivity to immunogenic MAMPs can be altered by the plant increasing/decreasing FLS2 expression. For instance, FLS2 expression increases upon cellular damage (Zhou et al., 2020). This allows plants to alter the proportion of suppression signals needed to buffer the presence of immunogenic MAMPs. This system would

allow plants to simultaneously accommodate community formation while monitoring for pathogens.

Our data may also explain how plants deal with the spatial distribution of bacterial colonies. Bacteria generally exist in homogeneous microcolonies and biofilms (Pandit et al., 2020). The dispersal of immune inducing MAMP proteins is likely limited by the presence of a complex commensal community. The ability of commensal community members to antagonize, modulate, and suppress MTI would allow plants to produce targeted defense at locations with high densities of bacteria that produce immune inducing MAMPs. Thus, I hypothesize that the production of immune suppressing, antagonizing, and modulating molecules by commensal communities allows plants to produce a targeted immune response around bacteria that produce immunogenic MAMPs.

My work has also set the stage for enhancing our understanding of the function of PRRs. First, we identify flg22 variants that differentially activate plant responses, which will allow us to better understand the complexity of MAMP recognition. It also presents an opportunity to understand how MAMPs can induce defense against pathogens without detrimental effects on plant growth. The fact that MAMPs provide class level bacterial information, is tantalizing. It's possible that plants use MAMPs to recognize the type of pathogen causing dysbiosis and reproducibly respond, re-sort, and re-assemble their communities. Supporting this view, glycans, which are non-proteinaceous MAMPs, were recently found to predict class and phylum information of the microbe it's derived from (Chaliha et al., 2018, Bojar et al., 2021). It remains to be seen if plant responses to taxonomically associated MAMPs are correlated or not. This would provide clues to the function and evolution of PRR receptors. Lastly, the finding that diverse bacteria harbor non-overlapping host immune suppression mechanisms reveals how

limited our knowledge of suppression mechanisms is. With the expanded complexity uncovered here, it becomes imperative to use our new understanding to identify and characterize emergent simplicity that can be used to predict and explain community formation observed in nature. This would allow us to improve the efficacy of microbial inoculants and hopefully sustainably increase crop yields.

REFERENCES

Bojar, D., Powers, R.K., Camacho, D.M., and Collins, J.J. (2021). Deep-Learning Resources for Studying Glycan-Mediated Host-Microbe Interactions. Cell Host & Microbe *29*.

Chaliha, C., Rugen, M.D., Field, R.A., and Kalita, E. (2018). Glycans as Modulators of Plant Defense Against Filamentous Pathogens.Frontiers in Plant Science *9*.

Pandit, A., Adholeya, A., Cahill, D., Brau, L., and Kochar, M. (2020). Microbial biofilms in nature: Unlocking their potential for agricultural applications. Journal of Applied Microbiology *129*, 199–211.

Zhou, F., Emonet, A., Dénervaud Tendon, V., Marhavy, P., Wu, D., Lahaye, T., and Geldner, N. (2020). Co-incidence of Damage and Microbial Patterns Controls Localized Immune Responses in Roots. Cell *180*, 440-453.e418.